

SEARCH REQUEST FORM

U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office

Requestor's

Name:

Khatol Shahman-shah

Serial

Number:

09747,521

Date:

11/20/01

Phone:

308-8896

Art Unit:

1645

Search Topic:

Room # 80-17

Mail Box #

8E12

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

① Please search claims 1-30 attached.

Full search including author (see bib sheet)
Keywords:

B. anthracis

LF (Lethal Factor Protein)

Immunogenic Composition

PA (Protective Protein)

see attached claims 1-30 and abstract

② Please do a Seq Search on

Seq Id # 1

Poly nucleotide

Full search and residue 610-2295 Seq Id # 1

Seq Id # 2

Amino acid

Full search and residue

1-735

175-735 Seq Id # 4

1-775

9-202

Seq Id

STIC

NOV 21 2001

RECEIVED

STAFF USE ONLY

Date completed: 12/3/01

Searcher:

Sheppane

Terminal time:

Elapsed time:

CPU time:

Total time:

Number of Searches:

Number of Databases:

Search Site

STIC

CM-I

Pre-S

Type of Search

N.A. Sequence

A.A. Sequence

Structure

Bibliographic

Vendors

IG Suite

STN

Dialog

APS

Geninfo

SDC

DARC/Questel

Other

THIS PAGE BLANK (USPTO)

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:49:24 ; Search time 55.64 Seconds
(without alignments)
-110.818 Million cell updates/sec

Title: US-09-747-521-2_COPY_9_282
Perfect score: 1404
Sequence: 1 KVISMSCLVTAITLSPVFI.....RDVLQLYAPEAFNYMDKFN 274

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA.*

- 1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/6C.COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/6D.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY*

Result No.	Score	Match	Length	DB	Description
------------	-------	-------	--------	----	-------------

-copy_9_282.ra

Page 2

Db	1	AGGHGDCVGHVKEKKNKDEENKQOEHLKIMHIYKIEVGEEAVKKEAE	60
Qy	86	KLEKVPDVLKMYKAIGKGIYVGDITKHISLEALSEDKKKIKDIYGDALLHEHYV	145
Db	61	KLEKVPDVLKMYKAIGKGIYVGDITKHISLEALSEDKKKIKDIYGDALLHEHYV	120
Qy	146	AKEGYEPVLVIOSEDVTEKALNYYEIGKILSDILSKINQPYQKFLDVLNTIKNA	205
Db	121	AKEGYEPVLVIOSEDVTEKALNYYEIGKILSDILSKINQPYQKFLDVLNTIKNA	180
Qy	206	SDSDGQDLLFTNQLKEHPTDFSEVFLEQNSNEVQVFAKAFAYIIEPQRDVLQYAPEA	265
Db	181	SDSDGQDLLFTNQLKEHPTDFSEVFLEQNSNEVQVFAKAFAYIIEPQRDVLQYAPEA	240
Qy	266	FNMYDKFNE 274	
Db	241	FNMYDKFNE 249	

RESULT 3

RESULT 1
US-08-021-601-6
; Sequence 6, Application US/08021601
; Patent No. 5591631
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Nichols, Peter J.
; APPLICANT: Arora, Naveen
; APPLICANT: Singh, Yogendra
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,601
; FILING DATE: 19930212
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-9880
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-021-601-6

Query Match 91.3%; Score 1282; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 6.6e-98;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGGHGDCVGHVKEKKNKDEENKQOEHLKIMHIYKIEVGEEAVKKEAE 85

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:49:25 ; Search time 55.64 Seconds
(Without alignments)
308.996 Million cell updates/sec

Title: US-09-747-521-4
3913
Sequence: 1 MKRRKVLIPALSTIIIVS.....TSTNGIKKILIFSKKGYEIG 764

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents_AA:*
2: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/BACKFILE1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3667	93.7	735	1	US-08-021-601-4
2	3667	93.7	735	1	US-08-082-849B-4
3	3667	93.7	735	5	PCT-US94-01624-4
4	3622.5	92.6	903	1	US-08-021-601-12
5	3622.5	92.6	903	1	US-08-082-849B-12
6	3622.5	92.6	903	5	PCT-US94-01624-12
7	3490	89.2	719	5	US-08-082-849B-31
8	3490	89.2	719	1	PCT-US94-01624-31
9	789.5	20.2	884	1	US-08-471-033-5
10	789.5	20.2	884	2	US-08-471-044-5
11	789.5	20.2	884	2	US-08-463-483A-5
12	789.5	20.2	884	2	US-08-471-046A-5
13	789.5	20.2	884	2	US-08-470-566B-5
14	789.5	20.2	884	3	US-08-469-334-5
15	789.5	20.2	884	3	US-09-300-529-5
16	789.5	20.2	1346	1	US-08-471-033-23
17	789.5	20.2	1346	2	US-08-471-044-23
18	789.5	20.2	1346	2	US-08-463-483A-23
19	789.5	20.2	1346	2	US-08-471-046A-23
20	789.5	20.2	1346	2	US-08-470-566B-23
21	789.5	20.2	1346	2	US-08-469-334-23
22	789.5	20.2	1346	3	US-09-300-529-23
23	787	20.1	881	4	US-08-960-780-32
24	787	20.1	881	4	US-09-073-898-32
25	777	19.9	1338	1	US-08-471-033-50
26	777	19.9	1338	2	US-08-471-044-50
27	777	19.9	1338	2	US-08-463-483A-50

28	777	19.9	1338	2	US-08-471-046A-50	Sequence 50, Appl
29	777	19.9	1338	2	US-08-470-566B-50	Sequence 50, Appl
30	777	19.9	1338	2	US-08-469-334-50	Sequence 50, Appl
31	777	19.9	1338	3	US-09-300-529-50	Sequence 50, Appl
32	776.5	19.8	784	4	US-09-371-913A-7	Sequence 7, Appl1
33	775	19.8	852	1	US-08-471-033-36	Sequence 36, Appl1
34	775	19.8	852	2	US-08-471-044-36	Sequence 36, Appl1
35	775	19.8	852	2	US-08-463-483A-36	Sequence 36, Appl1
36	775	19.8	852	2	US-08-470-566B-36	Sequence 36, Appl1
37	775	19.8	852	2	US-08-470-566B-36	Sequence 36, Appl1
38	775	19.8	852	2	US-08-469-334-36	Sequence 36, Appl1
39	775	19.8	852	3	US-09-300-529-36	Sequence 36, Appl1
40	766	19.6	834	1	US-08-471-033-21	Sequence 21, Appl1
41	766	19.6	834	2	US-08-471-044-21	Sequence 21, Appl1
42	766	19.6	834	2	US-08-463-483A-21	Sequence 21, Appl1
43	766	19.6	834	2	US-08-471-046A-21	Sequence 21, Appl1
44	766	19.6	834	2	US-08-470-566B-21	Sequence 21, Appl1
45	766	19.6	834	2	US-08-469-334-21	Sequence 21, Appl1

ALIGNMENTS

RESULT 1
US-08-021-601-4
Sequence 4, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Ancora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-4
Query Match 93.7%; Score 3667; DB 1; Length 735;
Best local Similarity 97.6%; Pred. No. 2.6e-261;
Matches 717; Conservative 2; Mismatches 16; Indels 0; Gaps 0;
30 EVQENRLNESSSGGLGTFSDINFPQAPMYVTSITGDLSIPSELENIPSENOYF 89

```

|||||
Db 1 EVKQENRLNLESESSQGLGYFSDLNFOAPMVYTSSTGDLSPSSELENIPESENOYF 60
QY 90 QSAIWSGFIKYKKSDEYFATFASADNHVTMWVDDQEVINKKASNSNRIKLEKGLYQIKIY 149
Db 61 QSAIWSGFIKYKKSDEYFATFASADNHVTMWVDDQEVINKKASNSNRIKLEKGLYQIKIY 120
QY 150 QRENTEKGLDFKLYWTSQNKKEVISSDNQLPELKOKSSNRKRSTSGAPYVDDDN 209
Db 121 QRENTEKGLDFKLYWTSQNKKEVISSDNQLPELKOKSSNRKRSTSGAPYVDDDN 180
QY 210 DGIPLSLEVEGYTVYVKKRFTLSPWISNIHEKKGLTYKSSPEKWSASDPYSDFEYV 269
Db 181 DGIPLSLEVEGYTVYVKKRFTLSPWISNIHEKKGLTYKSSPEKWSASDPYSDFEYV 240
QY 270 GRIDKNSPEARHPLVAAYPIYHVMENIILSKNEDOSTQNTDSETRISKNTSRSTHT 329
Db 241 GRIDKNSPEARHPLVAAYPIYHVMENIILSKNEDOSTQNTDSETRISKNTSRSTHT 300
QY 330 SEVHGNAEVHANTSTSRHTSEVHGNAEVAHAYADHSLSLAGEERTWAGLNTADTARL 389
Db 301 SEVHGNAEVAHANTSTSRHTSEVHGNAEVAHAYADHSLSLAGEERTWAGLNTADTARL 360
QY 390 NANIRYVNTGAPRIYVNLPTTSLVGLKNOTLATIKAKENQSLIAPNNYPSKMLADIA 449
Db 361 NANIRYVNTGAPRIYVNLPTTSLVGLKNOTLATIKAKENQSLIAPNNYPSKMLADIA 420
QY 450 LNAODFSSPTITMNYNOFLELEKTKQLRLDQYGYGNATYNFENGVRVDTGSNMSEV 509
Db 421 LNAODFSSPTITMNYNOFLELEKTKQLRLDQYGYGNATYNFENGVRVDTGSNMSEV 480
QY 510 LPQIETTARIIFNGKDLNVERIAAANPSDPLETTKPDMLTKALIAEGFNPNGNL 569
Db 481 LPQIETTARIIFNGKDLNVERIAAANPSDPLETTKPDMLTKALIAEGFNPNGNL 540
QY 570 QYOGKDITEFDENFDQOQSNIKNOLAELNATNITYVLDKIKLNKMNILIRDKRFHYDR 629
Db 541 QYOGKDITEFDENFDQOQSNIKNOLAELNATNITYVLDKIKLNKMNILIRDKRFHYDR 600
QY 630 NNIAVGADESVYKKAHREVINSSTEGLLINDKDIRKILSGYIVIEDTEGLKEVIYNDY 689
Db 601 NNIAVGADESVYKKAHREVINSSTEGLLINDKDIRKILSGYIVIEDTEGLKEVIYNDY 660
QY 660 DMLNSSLRODGKTFIDKRYNDKLPYISNPYKVNYAYATKENTIIINPSNGSTSTNG 749
Db 661 DMLNSSLRODGKTFIDKRYNDKLPYISNPYKVNYAYATKENTIIINPSNGSTSTNG 720
QY 750 IKKILIFSKKGEIG 764
Db 721 IKKILIFSKKGEIG 735

```

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-4

Query Match 93.7%; Score 3667; DB 1; Length 735;
Best Local Similarity 97.6%; Pred. No. 2.6e-261;
Matches 717; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

Db 30 EVKQENRLNLESESSQGLGYFSDLNFOAPMVYTSSTGDLSPSSELENIPESENOYF 89
Db 1 EVKQENRLNLESESSQGLGYFSDLNFOAPMVYTSSTGDLSPSSELENIPESENOYF 60
QY 90 QSAIWSGFIKYKKSDEYFATFASADNHVTMWVDDQEVINKKASNSNRIKLEKGLYQIKIY 149
Db 61 QSAIWSGFIKYKKSDEYFATFASADNHVTMWVDDQEVINKKASNSNRIKLEKGLYQIKIY 120
QY 150 QRENTEKGLDFKLYWTSQNKKEVISSDNQLPELKOKSSNRKRSTSGAPYVDDDN 209
Db 121 QRENTEKGLDFKLYWTSQNKKEVISSDNQLPELKOKSSNRKRSTSGAPYVDDDN 180
QY 210 DGIPLSLEVEGYTVYVKKRFTLSPWISNIHEKKGLTYKSSPEKWSASDPYSDFEYV 269
Db 181 DGIPLSLEVEGYTVYVKKRFTLSPWISNIHEKKGLTYKSSPEKWSASDPYSDFEYV 240
QY 270 GRIDKNSPEARHPLVAAYPIYHVMENIILSKNEDOSTQNTDSETRISKNTSRSTHT 329
Db 241 GRIDKNSPEARHPLVAAYPIYHVMENIILSKNEDOSTQNTDSETRISKNTSRSTHT 300
QY 330 SEVHGNAEVHANTSTSRHTSEVHGNAEVAHAYADHSLSLAGEERTWAGLNTADTARL 389
Db 301 SEVHGNAEVAHANTSTSRHTSEVHGNAEVAHAYADHSLSLAGEERTWAGLNTADTARL 360
QY 390 NANIRYVNTGAPRIYVNLPTTSLVGLKNOTLATIKAKENQSLIAPNNYPSKMLADIA 449
Db 361 NANIRYVNTGAPRIYVNLPTTSLVGLKNOTLATIKAKENQSLIAPNNYPSKMLADIA 420
QY 450 LNAODFSSPTITMNYNOFLELEKTKQLRLDQYGYGNATYNFENGVRVDTGSNMSEV 509
Db 421 LNAODFSSPTITMNYNOFLELEKTKQLRLDQYGYGNATYNFENGVRVDTGSNMSEV 480
QY 510 LPQIETTARIIFNGKDLNVERIAAANPSDPLETTKPDMLTKALIAEGFNPNGNL 569
Db 481 LPQIETTARIIFNGKDLNVERIAAANPSDPLETTKPDMLTKALIAEGFNPNGNL 540
QY 570 QYOGKDITEFDENFDQOQSNIKNOLAELNATNITYVLDKIKLNKMNILIRDKRFHYDR 629
Db 541 QYOGKDITEFDENFDQOQSNIKNOLAELNATNITYVLDKIKLNKMNILIRDKRFHYDR 600
QY 630 NNIAVGADESVYKKAHREVINSSTEGLLINDKDIRKILSGYIVIEDTEGLKEVIYNDY 689

```

```

RESULT 2
US-08-082-849B-4
: Sequence 4, Application US/08082849B
: Patent No. 5677274
: GENERAL INFORMATION:
: APPLICANT: Leppla, Stephen H.
: APPLICANT: Klimpel, Kurt R.
: APPLICANT: Arora, Naveen
: APPLICANT: Singh, Yogendra
: APPLICANT: Nichols, Peter J.
: TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
: NUMBER OF SEQUENCES: 35
: CORRESPONDENCE ADDRESS:
: ADDRESS: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:

```

Db 601 NNIAVGADESVKAEHREVINSSTEGLLNDKIDRKILSGYIEIEDTEGLKEVINDRY 660
QY 690 DMLNSSLRODGKTFIDFKRYNDKPLXISNPNKVVAVYATKNTIINPSENDTSTNG 749
Db 661 DMLNSSLRODGKTFIDFKRYNDKPLXISNPNKVVAVYATKNTIINPSENDTSTNG 720
QY 750 IKKILIFSKKGYEIG 764
Db 721 IKKILIFSKKGYEIG 735

RESULT 3
PCT-US94-01624-4
Sequence 4, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppia, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
STREET: Stewart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-4

Query Match 93.74; Score 3667; DB 5; Length 735;
Best Local Similarity 97.64; Pred. No. 2,6e-261;
Matches 717; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 30 EVKQENRLNSESSESSOGLGYYFSDNLFQAPMVVTSSTTGDLSIPSELENIPSENOYF 89
Db 1 EVKQENRLNSESSESSOGLGYYFSDNLFQAPMVVTSSTTGDLSIPSELENIPSENOYF 60
QY 90 OSAIWSGFIVKKSDEYTFATSDNHNVTMVDQDEVINKASNSKIRLEKGRLYQIKIY 149
Db 61 OSAIWSGFIVKKSDEYTFATSDNHNVTMVDQDEVINKASNSKIRLEKGRLYQIKIY 120
QY 150 QRENPTKGLDFKLYWDSQKKKEVYSSDNQLPELQKSSNSKRRKSTSGPYPPDDN 209
Db 121 QRENPTKGLDFKLYWDSQKKKEVYSSDNQLPELQKSSNSKRRKSTSGPYPPDDN 180
QY 210 DGIPDSLEVEGYTVDVKNKRTFLSPWISNIHEKKGLTRKYSPPKWSYASDPYSDFEAVT 269

Db 181 DGIPDSLEVEGYTVDVKNKRTFLSPWISNIHEKKGLTRKYSPPKWSYASDPYSDFEAVT 240
QY 270 GRIDKNVSPEARHPLVAAPYIVHVMENIILSKNEDOSTONTDESTRITSKNTSRTHT 329
Db 241 GRIDKNVSPEARHPLVAAPYIVHVMENIILSKNEDOSTONTDESTRITSKNTSRTHT 300
QY 330 SEVHGNAEVHANTSTRHTSEVHGNAEVHVAIDHSLSLGERTWAKTMLNADTARL 389
Db 301 SEVHGNAEVHASFDPDGSVAGFSNSSTVALDHSLSLGERTWAKTMLNADTARL 360
QY 390 NANIRYVNTGAPLYNVLPPTSLVGLKNOTLATIKAKENQLSQILAPNNYPSKNLAPIA 449
Db 361 NANIRYVNTGAPLYNVLPPTSLVGLKNOTLATIKAKENQLSQILAPNNYPSKNLAPIA 420
QY 450 LNAODDESSPTITMNYNOFLEKTKQLRLPTDOYGYNIATYNEBNGRVAVDTGSNNSEV 509
Db 421 LNAODDESSPTITMNYNOFLEKTKQLRLPTDOYGYNIATYNEBNGRVAVDTGSNNSEV 480
QY 510 LPOIETTARIIFNGKDLNVERRIAANVPSDPLETTKPDMTLEALIKAFGNPNGL 569
Db 481 LPOIETTARIIFNGKDLNVERRIAANVPSDPLETTKPDMTLEALIKAFGNPNGL 540
QY 570 QYQKIDITEFDNFDOQTSQNIKNQLAELNATNTIYVLDKIKLAKNNILIRDRKFHYDR 629
Db 541 QYQKIDITEFDNFDOQTSQNIKNQLAELNATNTIYVLDKIKLAKNNILIRDRKFHYDR 600
QY 630 NNIAVGADESVKAEHREVINSSTEGLLNDKIDRKILSGYIEIEDTEGLKEVINDRY 689
Db 601 NNIAVGADESVKAEHREVINSSTEGLLNDKIDRKILSGYIEIEDTEGLKEVINDRY 660
QY 690 DMLNSSLRODGKTFIDFKRYNDKPLXISNPNKVVAVYATKNTIINPSENDTSTNG 749
Db 661 DMLNSSLRODGKTFIDFKRYNDKPLXISNPNKVVAVYATKNTIINPSENDTSTNG 720
QY 750 IKKILIFSKKGYEIG 764
Db 721 IKKILIFSKKGYEIG 735

RESULT 4
US-08-021-601-12
Sequence 12, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppia, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratl, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414,057

TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
us-08-021-601-12

Query Match 92.6%; Score 3622.5; DB 1; Length 903;
Best Local Similarity 96.7%; Pred. No. 6.5e-258;

Matches 710; Conservative 3; Mismatches 18; Indels 3; Gaps 1;

QY 30 EYKQENRLNSESSESSQGLGYYFSDNLFQAPMVVTSSTTGDLSPSSLENIIPSENQYF 89
DB 1 EYKQENRLNSESSESSQGLGYYFSDNLFQAPMVVTSSTTGDLSPSSLENIIPSENQYF 60
QY 90 QSAIWSGFIKVKKSDDEYTPATSDAHNVHTMVDQEVINKASNSNKIRLEKGRLYQIKIY 149
DB 61 QSAIWSGFIKVKKSDDEYTPATSDAHNVHTMVDQEVINKASNSNKIRLEKGRLYQIKIY 120
QY 150 QRENPEKGLDFKLYWTDSONKKEVISSDNQLPELKOKSSNRKRKRSAGPTVPDRDN 209
DB 121 QRENPEKGLDFKLYWTDSONKKEVISSDNQLPELKOKSSNRKRKRSAGPTVPDRDN 180
QY 210 DGIPOSLEVEGYTVDYKKNRFTLSPWISNIHEKKGLTKYKSSPEKSTASDPYSDPEKYT 269
DB 181 DGIPOSLEVEGYTVDYKKNRFTLSPWISNIHEKKGLTKYKSSPEKSTASDPYSDPEKYT 240
QY 270 GRIDKNVSPPEARHPLVAAPIYHVDMENTILSKNEDOSTQNTDSEFRTISKNTSRTHT 329
DB 241 GRIDKNVSPPEARHPLVAAPIYHVDMENTILSKNEDOSTQNTDSEFRTISKNTSRTHT 300
QY 330 SEVHGNAEYHANTSTSRHTSEVHGNAEYHAVAIDHSLSLAGEERTWAGTADTARL 389
DB 301 SEVHGNAEYHANTSTSRHTSEVHGNAEYHAVAIDHSLSLAGEERTWAGTADTARL 360
QY 390 NANIRYVNTGTAPIYVNLPTTSLVLGKNOTLATIKAKENQLSQILAPNNYPSKNLAPIA 449
DB 361 NANIRYVNTGTAPIYVNLPTTSLVLGKNOTLATIKAKENQLSQILAPNNYPSKNLAPIA 420
QY 450 LNAODDFSSPTITMANTNOFLLEKTKQLRLDQVYGNATYVNFENGRRVVDGSMSEY 509
DB 421 LNAODDFSSPTITMANTNOFLLEKTKQLRLDQVYGNATYVNFENGRRVVDGSMSEY 480
QY 510 LPOIOTETARIIFENGKDLNVERRIAANVPSDLETTKPDMLKEALKIAPFNPENGNL 569
DB 481 LPOIOTETARIIFENGKDLNVERRIAANVPSDLETTKPDMLKEALKIAPFNPENGNL 540
QY 570 QYQKGDITEFDENFDQOSTQNIKNQDLAELNATNITYVLDKIKLMAKMNILIDKRRHYR 629
DB 541 QYQKGDITEFDENFDQOSTQNIKNQDLAELNATNITYVLDKIKLMAKMNILIDKRRHYR 600
QY 630 NNIAGADSVYKKAHREYINSSTEGLLINDKDIRKILSGIYVETEDTEGKEVINDRY 689
DB 601 NNIAGADSVYKKAHREYINSSTEGLLINDKDIRKILSGIYVETEDTEGKEVINDRY 660
QY 690 DMLNTSSLRODGKTFIDFKKYNDKLPYISNPYKVNVAVKENTIIIPSENGDSTNG 749
DB 661 DMLNTSSLRODGKTFIDFKKYNDKLPYISNPYKVNVAVKENTIIIPSENGDSTNG 720
QY 750 IKKIL---IFSCKG 760
DB 721 IKKILKKVYLGKKG 734

RESULT 5
US-08-082-849B-12
Sequence 12, Application US/08082849B
Patent No. 5677274

GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
TITLE OF INVENTION: Related Methods
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
us-08-082-849B-12

Query Match 92.6%; Score 3622.5; DB 1; Length 903;
Best Local Similarity 96.7%; Pred. No. 6.5e-258;
Matches 710; Conservative 3; Mismatches 18; Indels 3; Gaps 1;

QY 30 EYKQENRLNSESSESSQGLGYYFSDNLFQAPMVVTSSTTGDLSPSSLENIIPSENQYF 89
DB 1 EYKQENRLNSESSESSQGLGYYFSDNLFQAPMVVTSSTTGDLSPSSLENIIPSENQYF 60
QY 90 QSAIWSGFIKVKKSDDEYTPATSDAHNVHTMVDQEVINKASNSNKIRLEKGRLYQIKIY 149
DB 61 QSAIWSGFIKVKKSDDEYTPATSDAHNVHTMVDQEVINKASNSNKIRLEKGRLYQIKIY 120
QY 150 QRENPEKGLDFKLYWTDSONKKEVISSDNQLPELKOKSSNRKRKRSAGPTVPDRDN 209
DB 121 QRENPEKGLDFKLYWTDSONKKEVISSDNQLPELKOKSSNRKRKRSAGPTVPDRDN 180
QY 210 DGIPOSLEVEGYTVDYKKNRFTLSPWISNIHEKKGLTKYKSSPEKSTASDPYSDPEKYT 269
DB 181 DGIPOSLEVEGYTVDYKKNRFTLSPWISNIHEKKGLTKYKSSPEKSTASDPYSDPEKYT 240
QY 270 GRIDKNVSPPEARHPLVAAPIYHVDMENTILSKNEDOSTQNTDSEFRTISKNTSRTHT 329
DB 241 GRIDKNVSPPEARHPLVAAPIYHVDMENTILSKNEDOSTQNTDSEFRTISKNTSRTHT 300
QY 330 SEVHGNAEYHANTSTSRHTSEVHGNAEYHAVAIDHSLSLAGEERTWAGTADTARL 389
DB 301 SEVHGNAEYHANTSTSRHTSEVHGNAEYHAVAIDHSLSLAGEERTWAGTADTARL 360
QY 390 NANIRYVNTGTAPIYVNLPTTSLVLGKNOTLATIKAKENQLSQILAPNNYPSKNLAPIA 449

```

Db 361 NANIRYVNTGTADIVNVLPTTSLVGLGKQTLATIKAKENQLSQLIAPNNYPSKNLAPIA 420
QY 450 LNAODDESSPTITMNNYNOFLEKTKQLRLDQOVYGNINATYNEENRVRVDTGSNNSEV 509
Db 421 LNAODDESSPTITMNNYNOFLEKTKQLRLDQOVYGNINATYNEENRVRVDTGSNNSEV 480
QY 510 LPOIETTARIIFNGKDLNVERRIAIVNPSDPLETTKPPMTLKEALKIAGFNEPNCNL 569
Db 481 LPOIETTARIIFNGKDLNVERRIAIVNPSDPLETTKPPMTLKEALKIAGFNEPNCNL 540
QY 570 QYQKDTTEFDNFDDQTSQNIKNQALNATNITYVLDRKIKLAKNNILIRDRKFHYDR 629
Db 541 QYQKDTTEFDNFDDQTSQNIKNQALNATNITYVLDRKIKLAKNNILIRDRKFHYDR 600
QY 630 NNIAVGADESVYKAEHREVINSSTEGLLNIDDKIRKILSGYIEIDTEGLKEVINDRY 689
Db 601 NNIAVGADESVYKAEHREVINSSTEGLLNIDDKIRKILSGYIEIDTEGLKEVINDRY 660
QY 690 DMLNSSLRODGKTFIDFKKYNDKPLIYISNPYKVVNYAVTKENTIIINPSENGDTSTNG 749
Db 661 DMLNSSLRODGKTFIDFKKYNDKPLIYISNPYKVVNYAVTKENTIIINPSENGDTSTNG 720
QY 750 IKKIL--IFSCKG 760
Db 721 IKKILKKVVLGCKG 734

```

```

RESULT 6
PCT-US94-01624-12
: Sequence 12, Application PC/TUS9401624
: GENERAL INFORMATION:

```

```

: APPLICANT: Leppia, Stephen H.
: APPLICANT: Kimpel, Kurt R.
: APPLICANT: Airoa, Naveen
: APPLICANT: Singh, Yogendra
: APPLICANT: Nichols, Peter J.
: TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
: TITLE OF INVENTION: RELATED METHODS
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
: STREET: Steuart Street Tower, 20th Floor, One Market
: STREET: Plaza
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94105

```

```

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/01624
: FILING DATE: June 25, 1993

```

```

: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Weber, Kenneth A.
: REGISTRATION NUMBER: 31,677
: REFERENCE/DOCKET NUMBER: 15280-115
: TELEPHONE: (415) 543-9600
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 903 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US94-01624-12

```

```

Query Match 92.6%; Score 3622.5; DB 5; Length 903;
Best Local Similarity 96.7%; Pred. No. 6,5e-258;
Matches 710; Conservative 3; Mismatches 18; Indels 3; Gaps 1;
QY 30 EVKQENRLNSESSESSGGLGYRSDLNFOAPMYVTSSTGDLISPSSELENISENOYF 89
Db 1 EVKQENRLNSESSESSGGLGYRSDLNFOAPMYVTSSTGDLISPSSELENISENOYF 60
QY 90 QSAIWGFIKKKSDTEFTFATSADNHYTMWDDQEVINKASNSKIRLEKRLQIRKQY 149
Db 61 QSAIWGFIKKKSDTEFTFATSADNHYTMWDDQEVINKASNSKIRLEKRLQIRKQY 120
QY 150 QRENPTKGLDFKLYMTDSQNKKEVSSDNQLDELKOKSSNRKSTSGAPVPPRDN 209
Db 121 QRENPTKGLDFKLYMTDSQNKKEVSSDNQLDELKOKSSNRKSTSGAPVPPRDN 180
QY 210 DGIPDSLEVEGYIVDVKNKRTFLSPWISNHEKGLTKYKSPKWSIASOPYSDFEYVT 269
Db 181 DGIPDSLEVEGYIVDVKNKRTFLSPWISNHEKGLTKYKSPKWSIASOPYSDFEYVT 240
QY 270 GRIDKNVSPKARHPLVAAYPIVHYDMENIILSKNEDOSTQMTDSEPTISKNTSTRTHT 329
Db 241 GRIDKNVSPKARHPLVAAYPIVHYDMENIILSKNEDOSTQMTDSEPTISKNTSTRTHT 300
QY 330 SEVHGNAEVRHANTSTRTHTSEVHGNAEVRHAVAIDHSLSLAGERTMAETMGLNADTARL 389
Db 301 SEVHGNAEVRHANTSTRTHTSEVHGNAEVRHAVAIDHSLSLAGERTMAETMGLNADTARL 360
QY 390 NANIRYVNTGTADIVNVLPTTSLVGLGKQTLATIKAKENQLSQLIAPNNYPSKNLAPIA 449
Db 361 NANIRYVNTGTADIVNVLPTTSLVGLGKQTLATIKAKENQLSQLIAPNNYPSKNLAPIA 420
QY 450 LNAODDESSPTITMNNYNOFLEKTKQLRLDQOVYGNINATYNEENRVRVDTGSNNSEV 509
Db 421 LNAODDESSPTITMNNYNOFLEKTKQLRLDQOVYGNINATYNEENRVRVDTGSNNSEV 480
QY 510 LPOIETTARIIFNGKDLNVERRIAIVNPSDPLETTKPPMTLKEALKIAGFNEPNCNL 569
Db 481 LPOIETTARIIFNGKDLNVERRIAIVNPSDPLETTKPPMTLKEALKIAGFNEPNCNL 540
QY 570 QYQKDTTEFDNFDDQTSQNIKNQALNATNITYVLDRKIKLAKNNILIRDRKFHYDR 629
Db 541 QYQKDTTEFDNFDDQTSQNIKNQALNATNITYVLDRKIKLAKNNILIRDRKFHYDR 600
QY 630 NNIAVGADESVYKAEHREVINSSTEGLLNIDDKIRKILSGYIEIDTEGLKEVINDRY 689
Db 601 NNIAVGADESVYKAEHREVINSSTEGLLNIDDKIRKILSGYIEIDTEGLKEVINDRY 660
QY 690 DMLNSSLRODGKTFIDFKKYNDKPLIYISNPYKVVNYAVTKENTIIINPSENGDTSTNG 749
Db 661 DMLNSSLRODGKTFIDFKKYNDKPLIYISNPYKVVNYAVTKENTIIINPSENGDTSTNG 720
QY 750 IKKIL--IFSCKG 760
Db 721 IKKILKKVVLGCKG 734

```

```

RESULT 7
US-08-082-849B-31
: Sequence 31, Application US/08082849B
: Patent No. 5677274
: GENERAL INFORMATION:
: APPLICANT: Leppia, Stephen H.
: APPLICANT: Kimpel, Kurt R.
: APPLICANT: Airoa, Naveen
: APPLICANT: Singh, Yogendra
: APPLICANT: Nichols, Peter J.
: TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
: TITLE OF INVENTION: Related Methods
: NUMBER OF SEQUENCES: 35
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor

```

CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentlin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-31

Query Match 89.2%; Score 3490; DB 1; Length 719;
Best Local Similarity 93.1%; Pred. No. 2.5e-248;
Matches 688; Conservative 6; Mismatches 21; Indels 24; Gaps 2;

OY 30 EVKQENRLNSESSESSQGLGYFSDLNFOAPMVYTSSTGDLSPSSELENIPSENOYF 89
DB 1 EVKQENRLNSESSESSQGLGYFSDLNFOAPMVYTSSTGDLSPSSELENIPSENOYF 60
OY 90 QSAIWSGFIKVKKSDEYFATSDNHVTMWVDDDEVINKASNSNKRIRLEKGRLYQIKIOY 149
DB 61 QSAIWSGFIKVKKSDEYFATSDNHVTMWVDDDEVINKASNSNKRIRLEKGRLYQIKIOY 120
OY 150 QRENPTTEKGLDFKLYWDSQNKKEVISSDNLQLPCLKSSNS---RRKRSTAGPTVP 205
DB 121 QRENPTTEKGLDFKLYWDSQNKKEVISSDNLQLPCLKSSNSNTATIMQRCNFTLQGPVP 180
OY 206 DRDNCIGDPSLEVEGYTVDVKNKRTEFLSPRISNIHEKKGITRYKSSPEKWSASPDYDF 265
DB 181 DRDNCIGDPSLEVEGYTVDVKNKRTEFLSPRISNIHEKKGITRYKSSPEKWSASPDYDF 240
OY 266 EKVYGRIDKNVSPERHPLVAAVPIVHVDMEIILSKNEDOSTOITDSETRTISKNTSTS 325
DB 241 EKVYGRIDKNVSPERHPLVAAVPIVHVDMEIILSKNEDOSTOITDSETRTISKNTSTS 300
OY 326 RTHSEVHGNAEVAHANSTSRTHSEVHGNAEVAHAVALIDHSLSLAGEPTMAETMGLNTAD 385
DB 301 RTHSEVHGNAEVAHANSTSRTHSEVHGNAEVAHAVALIDHSLSLAGEPTMAETMGLNTAD 360
OY 386 TARLANRIRYNTGAPTYNVLPPTSLVIGNQTLATIKAKENOSQILAPNNYPSKNL 445
DB 361 TARLANRIRYNTGAPTYNVLPPTSLVIGNQTLATIKAKENOSQILAPNNYPSKNL 420
OY 446 APALNAODDSPTITANNYQFLELEKTKQLRLDTPQVGNIAIYNFENGSRVAVDTGSN 505
DB 421 APALNAODDSPTITANNYQFLELEKTKQLRLDTPQVGNIAIYNFENGSRVAVDTGSN 460
OY 506 WSEVLPQIOETTARITNGKDLNVERIAAVNPSDPLETTKPDMTLKEALKIAFGNEP 565
DB 461 WSEVLPQIOETTARITNGKDLNVERIAAVNPSDPLETTKPDMTLKEALKIAFGNEP 520
OY 566 NGNIQYQKDLTEPFDNFDOQTSQNIKNQDLAELNATNTIYVLDIRKLAKNMILIRDRKF 625

DB 521 NGNIQYQKDLTEPFDNFDOQTSQNIKNQDLAELNATNTIYVLDIRKLAKNMILIRDRKF 580
OY 626 HYDRNNIAGADESVYKKAHREVINSTEGGLNLNDKIDRKILSGYIETDETEGLKEVI 685
DB 581 HYDRNNIAGADESVYKKAHREVINSTEGGLNLNDKIDRKILSGYIETDETEGLKEVI 640
OY 666 NDRYMLNLISSLRQDGKTFIDFKKYNDKLPYISPNPKVNVYAVTKENTIIINPSENDT 745
DB 641 NDRYMLNLISSLRQDGKTFIDFKKYNDKLPYISPNPKVNVYAVTKENTIIINPSENDT 700
OY 746 STNGIKKILIFSKKGIEIG 764
DB 701 STNGIKKILIFSKKGIEIG 719

RESULT 8
PCT-US94-01624-31
Sequence 31, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Lepola, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arota, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
STREET: Stewart Street Tower, 20th Floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentlin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-31

Query Match 89.2%; Score 3490; DB 5; Length 719;
Best Local Similarity 93.1%; Pred. No. 2.5e-248;
Matches 688; Conservative 6; Mismatches 21; Indels 24; Gaps 2;

OY 30 EVKQENRLNSESSESSQGLGYFSDLNFOAPMVYTSSTGDLSPSSELENIPSENOYF 89
DB 1 EVKQENRLNSESSESSQGLGYFSDLNFOAPMVYTSSTGDLSPSSELENIPSENOYF 60
OY 90 QSAIWSGFIKVKKSDEYFATSDNHVTMWVDDDEVINKASNSNKRIRLEKGRLYQIKIOY 149
DB 61 QSAIWSGFIKVKKSDEYFATSDNHVTMWVDDDEVINKASNSNKRIRLEKGRLYQIKIOY 120


```

OY 150 QRENPTKGLDFKLYWTDSONKKEVISSDNLOLPELKOKSSNS-RRKRSTSGPTVP 205
    |||||
DB 121 QRENPTKGLDFKLYWTDSONKKEVISSDNLOLPELKOKSSNSRTATIMOGNFGPTVP 180
    |||||
OY 206 DRNDGIPDSLEVEGYTVVKKRRTFLSPWISINHEKKGLTKYKSSPEKSTASDPYSDF 265
    |||||
DB 181 DRNDGIPDSLEVEGYTVVKKRRTFLSPWISINHEKKGLTKYKSSPEKSTASDPYSDF 240
    |||||
OY 266 EKTGRIDKNVSPKARHPLAAAPYIVHVDMENIILSKNEOSTONTDSETRTISKNTS 325
    |||||
DB 241 EKTGRIDKNVSPKARHPLAAAPYIVHVDMENIILSKNEOSTONTDSETRTISKNTS 300
    |||||
OY 326 RHTSEVHGNAEVAHANTSTSRHTSEVHGNAEVAHVAIDHSLSLAGERTAEETGLTAD 385
    |||||
DB 301 RHTSEVHGNAEVAHANTSTSRHTSEVHGNAEVAHVAIDHSLSLAGERTAEETGLTAD 360
    |||||
OY 386 TARLNANIRVYNTGTAPYIVNLPPTSLVLGKNOTLATIKAKENOLSOILAPNNYPSKNL 445
    |||||
DB 361 TARLNANIRVYNTGTAPYIVNLPPTSLVLGKNOTLATIKAKENOLSOILAPNNYPSKNL 420
    |||||
OY 446 APALNAODFSSPTITMANNOLFELEKTKQLDLDQVYGNATATYFENGRRVDTGSN 505
    |||||
DB 421 APALNAODFSSPTITMANNOLFELEKTKQLDLDQVYGNATATYFENGRRVDTGSN 460
    |||||
OY 506 WSEVLPOIOETTARIIFNGKDLNVERRIAANVPSDELTTKPDMLKEALKAFAFGENP 565
    |||||
DB 461 WSEVLPOIOETTARIIFNGKDLNVERRIAANVPSDELTTKPDMLKEALKAFAFGENP 520
    |||||
OY 566 NGMLQYOGKDIETEPDEFNDOOTSQNIKNOLAEINATNIYVLDKIKLAKMMLIRDRF 625
    |||||
DB 521 NGMLQYOGKDIETEPDEFNDOOTSQNIKNOLAEINATNIYVLDKIKLAKMMLIRDRF 580
    |||||
OY 626 HYRNINAVGADSVKKEAREVINSSTEGLLINIDKIKLISGYVLEEDPBGKEVI 685
    |||||
DB 581 HYRNINAVGADSVKKEAREVINSSTEGLLINIDKIKLISGYVLEEDPBGKEVI 640
    |||||
OY 686 NDREYDMLNISLRODKTFFDKFKYNDKLYISNPKYKNVAVTKENTILINSENGDT 745
    |||||
DB 641 NDREYDMLNISLRODKTFFDKFKYNDKLYISNPKYKNVAVTKENTILINSENGDT 700
    |||||
OY 746 STNGIKKILIFSKKYEIG 764
    |||||
DB 701 STNGIKKILIFSKKYEIG 719
    |||||

RESULT 9
US-08-471-033-5
: Sequence 5, Application us/08471033
: Patent No. 5770696
: GENERAL INFORMATION:
: APPLICANT: Warren, Gregory M
: APPLICANT: Koziel, Michael G
: APPLICANT: Mullins, Martha A
: APPLICANT: Nye, Gordon J
: APPLICANT: Carr, Brian
: APPLICANT: Desai, Nalin M
: APPLICANT: Kostichka, N Kristy
: APPLICANT: Duck, Nicholas B
: APPLICANT: Estruch, Juan J
: TITLE OF INVENTION: NO. 5770696el Pesticidal Proteins and Strains
: NUMBER OF SEQUENCES: 50
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CIBA-GEIGY Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30B

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/471, 033
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/314, 594
: FILING DATE: 09-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/218, 018
: FILING DATE: 23-MAR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/037, 057
: FILING DATE: 25-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Pace, Gary M.
: REGISTRATION NUMBER: P-40,403
: REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8582
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 884 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-471-033-5

```

```

Query Match 20.2%; Score 789.5; DB 1; Length 884;
Best Local Similarity 29.4%; Pred. No. 8.3e-50;
Matches 253; Conservative 134; Mismatches 309; Indels 165; Gaps 34;

```

```

OY 1 MKKKKVLPLMALSTILVSS--TGNLEVIOAEV-----QENRLNBSSSSGLLG 51
    ||| : : : : : ||| : : : : : ||| : : : : : |||
DB 1 MKNNKKRLASVYCTLLAPMFLNGNVAAYADSKTNOISTTKNQ---QKEMDRKGLG 57
    ||| : : : : : ||| : : : : : ||| : : : : : |||
OY 52 YFSDLNFAQPVVYVSTTGDSIPSSLEEN--IPSENOYFOSAIWGSFIVKKSDEYFA 109
    ||| : : : : : ||| : : : : : ||| : : : : : |||
DB 58 YFGKDE-SNLTMPAPRDSSTLIYDQOTANKLKKQOETOSIWIILQISKETGDTFN 116
    ||| : : : : : ||| : : : : : ||| : : : : : |||
OY 110 TSADNHYTMVVDQEVINKASNSNKRIRLENGRLYQIRYOARENPTKGLD-----FKL 163
    ||| : : : : : ||| : : : : : ||| : : : : : |||
DB 117 LSEDEQALIELNGKIISNKKKEQVHLEKGLVPIKIEYQSD--TFPNIDSKYFKELKL 174
    ||| : : : : : ||| : : : : : ||| : : : : : |||
OY 164 YWTDSONKKEVYSSDNLOLPELKOKSS-----NSRKRSTSGAPVPPROND 210
    ||| : : : : : ||| : : : : : ||| : : : : : |||
DB 175 FKIDSONQPOOVQODELRNPEFNKESQEFLLAKPSKINLFTQKKREITDED--TDPDGD 231
    ||| : : : : : ||| : : : : : ||| : : : : : |||
OY 211 GTPDSLEVEGYTVVKKRRTFLSPWISINHEKKGLTKYKSSPEKSTASDPYSDEKVTG 270
    ||| : : : : : ||| : : : : : ||| : : : : : |||
DB 232 STPLMEENGTYT---QNRIVAKWDSL-ASKGYTFVSNPLSHVGDYDYDERAAR 286
    ||| : : : : : ||| : : : : : ||| : : : : : |||
OY 271 RIDKNVSPKARHPLAAAPYIVHVDMENIILSKNEOSTONTDSETRTISKNTSRTHTS 330
    ||| : : : : : ||| : : : : : ||| : : : : : |||
DB 287 DLDLSNKKETFPNPLVAAFPSVNVSMKVIILSPNNLS----- 323
    ||| : : : : : ||| : : : : : ||| : : : : : |||
OY 331 EVHGNAEVAHANTSTSRHTSEVHGNAEYH-----AVAIHSLSLAGERTAEETG-- 380
    ||| : : : : : ||| : : : : : ||| : : : : : |||
DB 324 ---NSVSHSSTMSYNTGEGASVEAGIGPKISFGVSVNQSHSETVAOE--WQTSAGNT 378
    ||| : : : : : ||| : : : : : ||| : : : : : |||
OY 381 --LNTADTARLNANIRVYNTGTAPYIVNLPPTSLVLGKNOTLATIKAKENOLSOILAPNN 438
    ||| : : : : : ||| : : : : : ||| : : : : : |||
DB 379 SQFN TASAGYLNANVNRNNGTGAIYDKPPTSVL--NNDITATITIKSNSTALNISBGE 437
    ||| : : : : : ||| : : : : : ||| : : : : : |||
OY 439 YPSKNLAPALNAODFSSPTITMANNOLFELEKTKQLRDLTDQVYGNATATYFENGRRV 498
    ||| : : : : : ||| : : : : : ||| : : : : : |||
DB 438 SYPKKGNGALITSMDDFNHPITLNNKKOYDNLNNKPMLETNQDGC---VYIKIDPHG 494
    ||| : : : : : ||| : : : : : ||| : : : : : |||
OY 499 RVDTGSMWSEVLPOIOETTARIIFNGKDLNVERRIAANVPSDELTTKPDMLKEALKI 558
    ||| : : : : : ||| : : : : : ||| : : : : : |||
DB 495 NIVYGGWNGVIOQIKAKTASIIIVDGE-RVAERKVAAKDYENEDKXT-PSLTLEKDALKL 552
    ||| : : : : : ||| : : : : : ||| : : : : : |||

```


RESULT 11
US-08-463-483A-5
Sequence 5, Application US/08463483A
Patent No. 5849870
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozielec, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalin M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5849870el Pesticidal proteins and strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-463-483A-5

Query Match 20.2%; Score 789.5; DB 2; Length 884;
Best Local Similarity 29.4%; Pred. No. 8.3e-50;
Matches 253; Conservative 134; Mismatches 309; Indels 165; Gaps 34;

QY 1 MKRKRVILPLMALSTIIYSS--TGNLEVIQAEV-----QENRLNNESSSSSGGLIGY 51
DB 1 MKNNKKLASVYCTLLAPMLNGNNAVYADSKTNOISTQKNQ---QKEMDRKGLGY 57
QY 52 YFSLNFOAPVWVSTTGDSIPSSSELEN--IPSENOYFOSATWSGFTYKKSDEYTF 109
DB 58 YFKKDF-SNLTJFAPRDSLTLLIDQOTANKLLKKQOEYOSIMWIGLQSKETGDFTFN 116
QY 110 TSADNHYTMVVDQEVINKASNSNKRIRLEKGRLYQIRIQYORENPTEKGLD-----FKL 163
DB 117 LSEDEQAIIEINGKIISNKKGEKVHLEKCKLVPIKIEYQSD--TKFNIDSKFFKELKL 174

QY 164 YWTDSONKKEVISSDNLQLELKOKSS-----NSRKRSTSGAPTVPRDND 210
DB 175 FKIDSONQPOQVOQDDELBNPEFNKKESEFELAKPSKINLFTQKKKRRIIDD--TPTDGD 231
QY 211 GIPDSLEVEGYTDVYKKNRFLSPWISNIEKKGLTYKKSPEKWSASDPYDFEYV 270
DB 232 SIPDLMEENGTYI---QNRIVAKWDDSL-ASKGYTFKFSNPLESHTVGPYDYDEKAR 286
QY 271 RIDKNVPEARHPVLAAYPIYVDMENIILSKNDOSTQMTDSTRTISKNTSRHTS 330
DB 287 DLDLSNAKETFPNPLVAAPSVNMEKVIISPENLS----- 323
QY 331 EVHGNAEVHANTSTSRHTSEVHGNAEVH-----AVAIDSLSLAGERTAEYWG-- 380
DB 324 ---NSVESHSTNWSYNTGASVEACIGPGISFGVSANQHSLEYAQE--WETSIGNT 378
QY 381 --LNTADTARILANIRYVNGTAPIYVLPPTSVLCKNOTLATIKKENOLAPNN 438
DB 379 SQFWTASAGYLANAVRYNNGTGAIDYVKPPTSEVL--NNDTIATITAKSNSTALNISPE 437
QY 439 YYPEKNLAPALANODPSSPTITMNNQFLEKTKQLRDTQVYGNATYVFNENGRV 498
DB 438 SYPKKGNGIAITSMDDFNHPITLNNKYDNNLNNKPMLETQDTG---VYIKDTHG 494
QY 499 RVDTGSNMSEVLPQIOETTARIIFNGKDLNVEBRIAIVNSDPLETTKPDMLKEALKI 558
DB 495 NIYTGGENGYIQIKAKTASIIYDDGE-RVAEKRVAAKYENEDBT-PSLTIKDLKL 552
QY 559 AF--GFNEPNGNLOYGKDIYEFDF--NFDQTSQNIKNQIAEL-----NATNIYVLDK 609
DB 553 SYPEIKETIEGLLYKKNPIYESSVMYTLDBENTAKETKQLNDTTGKFQVSHLYDV--- 609
QY 610 IKLNKANNILIRKRFHYDRNNINAVGADESYYKAAHEVINSSTEG-----LL 657
DB 610 -KLTPKMNVTIK-LSTLYDN--AESNDNSIGKWTNNIYSGNGNKKQYSSNPNANLT 664
QY 658 LIND-----KDIRKILSGYIVLEIDFE-----GLKEYINDRYDMLN---- 693
DB 665 LINTDAOEKLNNKRDYIISLYKSEKNOCETTINGELIYPTTKVYNNKQYKRLDIIAH 724
QY 694 -----ISSLRQDGTFLDFKRYNDKLPXYISNPNKYVAVYTKENTIIIPNSGDT-S 746
DB 725 NIKSNPISSLH-----IKTNDETITLFWDDISI--TDVASIKPEN--LTDSEIKQIYS 772
QY 747 TNGIK---KLIIFSKKGYEIG 764
DB 773 RYGIKLEDGILIDRKKGIIHG 793

RESULT 12
US-08-471-046A-5
Sequence 5, Application US/08471046A
Patent No. 5866326
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozielec, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalin M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: Method for Isolating Vegetative Insecticidal
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5866326artis Corporation
STREET: 3054 Cornwalis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,046A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-471-046A-5

Query Match      20.2%; Score 789.5; DB 2; Length 884;
Best Local Similarity 29.4%; Pred. No. 8.3e-50;
Matches 253; Conservative 134; Mismatches 309; Indels 165; Gaps 34;

QY 1 MKKRVLLPMLALSTLVSS--TGNLEVIQAEVK-----OENRLNSESSESSGGLGY 51
DB 1 MNMKKKLSVYTCILALMFLNGVNAVYADSKTQISTTQNO---OKEMDRGGLGY 57
QY 52 YPSDLNFOAPMYVSTSTGDLSPSSSELEN--IPSENQYFQSAIWSGFTKVKSDPYTPA 109
DB 58 YPKGKDF-SNLMPFAPTRBSTLIYDQGTANKLDDKKQGEYQSIKMTGLQSGSETGDTFN 116
QY 110 TSDADNHYTMVDDQEVYINKASNSNKIRLEKGRLYQIKIORYENPTEKGLD-----FKL 163
DB 117 LSEDEQALIEINGKILISNKGKQVVAHLEKGLVPIKIKIYQSD--TKFMDSKTEKELTL 174
QY 164 YMTDSQNKKEVISDNLOLPBLKOKS-----NSRKKRSTSGPTVPRDND 210
DB 175 FKIDSONQOQOQDELRLRPFENKKESEFLAKPSKINLFTQKMKREIDED---IDTDD 231
QY 211 GIPDLSLEVGYVVDVKNKRFSPWISNTHKKGTLTKYKSSPEKSTASDPYSDEKVTG 270
DB 232 SIPDLMEENGTYI---QRIAVKMDSD-ASKGYTKFVSNPLESHVTGDDPYTDEKKAAR 286
QY 271 RDKDNVSPARHPVAAYIVYHVDENITLSKNEOSTONTSEPTTKSTKSTHTS 330
DB 287 DLDLSNAKETFNPVLAAPFSVNVSMKYLSPNENLS----- 323
QY 331 EVHGNNAEVANTSTRTSHSEVHGNNAEVH-----AVAIHDSLAGEPTMAETMG-- 380
DB 324 ---NSVESHSSTNMWSTTNEGASVEAGICPKGISFGVSVNYOHSFVAOE--WGSTGTNT 378
QY 381 --LNTADTARLANIRVYVTGTAPYIVNLPPTSLVLGKNOTLATIKAKENQLSQILAPNN 438
DB 379 SQPNTASAGYLVANVRYNNVNGTGAIVDKVPTTISFVL--NNDTIATITAKSNSALANISPOE 437
```

```
QY 439 YPSPKNLAPIALMADDESSPTITMNVNOFLELEKTKOLRLDTPDOYGNIAIYFNENGRV 498
DB 438 SYPKKGONCIAITSMDFNSHPTLTKKQVDNLNNKKPMMLEETNOTDQ---VYKIKDTHG 494
QY 499 RYDTSQNSNNEVLPQIOETTARIIFNGKDLNLYERRIAVNPSPDLPTTKPKDMLKEALKI 558
DB 495 NIVTGGEMNGVIOQIKAKTASITIVDGE--RAVEKRVAAKDENPEDKT--PSLTLDALKL 552
QY 559 AF--GFNEPENGMLQYQKDKITEFDF--NEPDQTSQNIKNQLAEL-----NATIVYLDK 609
DB 553 SYPDEKEKEIEGLLYKKNRPYESSWVTYIDENTATAEVYKQLMDITGCKRQVSHLDV--- 609
QY 610 IKLAKMNLIPDKRPHYDRNNIAVGADESVYKAEHREYINSSTEG-----LL 657
DB 610 -KLTPKMNVTIK-LSILDYD--AESNDNSIGKMTNTNIVSGNNGKQYSSNPDANLT 664
QY 658 LVID-----KDIRKILSGYIYIEDTE-----GLKEVINDRDMEN----- 693
DB 665 LNTDQOEKLNKRNDYIISLYMKSEKNTQCEITIDGEIYPTTKTYVNVKNDYKRLDIIAH 724
QY 694 -----ISSLRQDCKFTIDFKYNDKLPYISNPKYKVVAVYKENTLINPSENGDT-S 746
DB 725 NIKSNPISLSLH-----IKTNDIEITLFWMDISI--TDVASIKPEN--LTPSEIKQIYS 772
QY 747 TNGIK--KILFSSKGYEIG 764
DB 773 RYGIKLEDDGILDKKGHNG 793

RESULT 13
US-08-470-566B-5
Sequence 5, Application US/08470566B
Patent No. 5872212
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Koslitchka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estrich, Juan J
TITLE OF INVENTION: No. 5872212a1 Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5872212a1 Pesticidal Proteins and Strains
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,566B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
```

```

: : ATTORNEY/AGENT INFORMATION:
: : NAME: Meigs, J. Timothy
: : REGISTRATION NUMBER: 38,241
: : REFERENCE/DOCKET NUMBER: GC16195/CIP3/DIV4 - SOLV4
: : TELECOMMUNICATION INFORMATION:
: : TELEPHONE: 919-541-8587
: : TELEFAX: 919-541-8689
: : INFORMATION FOR SEQ ID NO: 5:
: : SEQUENCE CHARACTERISTICS:
: : LENGTH: 884 amino acids
: : TYPE: amino acid
: : TOPOLOGY: linear
: : MOLECULE TYPE: protein
: : US-08-470-5668-5

Query Match 20.2%; Score 789.5; DB 2; Length 884;
Best Local Similarity 29.4%; Pred. No. 8.3e-50;
Matches 253; Conservative 134; Mismatches 309; Indels 165; Gaps 34

QY 1 MKKKRVLLPLALSTILVSS--TGNLEVIQAEVK-----QENRLNSESSSOGLLGY 51
Dd 1 MKNNKKKLASVYTCGLLAPMFLNGVAVNVAADSKTNQISTGKNQ--QKENDRGGLGY 57
QY 52 YPSDLNFOAPNVVYSTTGDDLSIPSELEN--IPSENOYFQSAINSGLIKVKKSDRYFA 109
Dd 58 YFKGKDF--SNLTMFAPTRDSTLIYDQQTANKLLDKKQOEYQSIIRYIGLSQKSTGDFPN 116
QY 110 TSAADHVTMMVDDQEVINKKASNSNKRIRLEKGRLOYKIDQYQENPTEKLD-----FTL 163
Dd 117 LSEDAQIIEINGKILISNKGKQYVHLEKGLVPLKIEYQSD--TKFNIDSKTFRELTL 174
QY 164 YWTDSONKREVISSDNLQPLPELKQKS-----NSRKKRSTSGPTVPRDND 210
Dd 175 FRIDSONQPOQVQOQDELNRNPEFNKESQEFIAKPSKINLFQKMKREIDED--TDTGCD 231
QY 211 GIPDSLEVEGTVDVYKKNRTFLSPWISNHEKKGLTKYKSPKMSASTADPISDEPKYVG 270
Dd 232 SIPDLMENSGYTI-----QNRILAVKDDSL-ASKGYTKFVSNPLESTHYVDPPYDEKAR 286
QY 271 RIDKRVSPEARPLVAAAPYIVHDMENILISKNDQSTQNTDSEPTTISKNTSTRTHS 330
Dd 287 DIDLSSNAKETNPVLAAPSPVANSHEKYLSPENELSS----- 323
QY 331 EYVHGAEEYHANTSTSRTHTEYVHGAEEV-----AAVIDHSISLAGERTAETMG-- 380
Dd 334 ---NVEGHSISSTNMSTYNTNEGASVEAGIQPKGISFGVSYNYQHSSEVAAE--WGISTGNT 378
QY 361 --LNTADTARLANIRIYVNTGTAPRIYVNLPTTSLVGLKNQTLATIKAKENQSLQILAPNN 438
Dd 379 SQFNTASAGYLIAANVRVNNVGTGAIYDVAKPTTSPVL--NNDTIATITAKSNSALTNISPEE 437
QY 439 YTPSKNLIAPIALNADDDPSPTITMKNYNQFLELEKTKQLRDTDDQYGGIATINFNENGRV 498
Dd 438 STPKKGQNGCIATSDDDFNFSHPITLTKKQVNDLNNKPMLEPTFNQDQ---YKIKIDTGG 494
QY 499 RVDTSQSNNSEVLPQIOETARLITFNQKQDNLNVERIRIAANPSPDLETPTKPDMLLKALKI 558
Dd 495 NIVYTGEGMNGVYIQQIKAKTASIIYDDGE--RVAEKRAAADYENPEPKT--PSLTLKALAL 552
QY 559 AF--GFNEPENGILQYQKDIITEFDF--NFDDQOTSONIKQNLQAEI-----NATNIYTVLDK 609
Dd 553 SYPDEIKETIEGLLYKKNKRIYESSVYTYIDENMTAKREVTKQLNDTGGKFKDVSILYDV-- 609
QY 610 IRLNKKMNLILPDKRFHYDRNMIIVAGADDSYVKEAREVINSTBEG-----LL 657
Dd 610 -RLTPKMANVTK-LSLTDN---AESNDNSIGKMTNTNIVSGGNKKQYSSNPNPANTL 664
QY 658 LIID-----KQIRKILSGYIVEIEDTE-----GLKEVINDRYDMLN----- 693
Dd 665 LNTDQAEKLNKRDYIISLYMKSEKTKQCEIITIDGCIYPTTKTYVANKDNYKRLDIIAH 724
QY 694 -----ISSLRDQGTETIDFKKYNKDLPLYSNPKNYVAVYATKENTIIINPSENGDT--S 746

```

```

Db 725 NIKSNPISLH-----IKTNDLITLFWDDISI-IDVASIKREN--LINDSEIKOIVS 772
QY 747 TNGIK--KILIFSKGYEIG 764
   ||| ||| | |
Db 773 RYGIKLEDGILLDKKGIIHYG 793

RESULT 14
US-08-469-334-5
: Sequence 5, Application US/08469334
: Patent No. 5990383
:
GENERAL INFORMATION:
: APPLICANT: Warren, Gregory W
: APPLICANT: Koziel, Michael G
: APPLICANT: Mullins, Martha A
: APPLICANT: Nye, Gordon J
: APPLICANT: Carr, Brian
: APPLICANT: Desai, Malini M
: APPLICANT: Kostichka, N. Kristly
: APPLICANT: Duck, Nicholas J
: APPLICANT: Estruch, Juan J
: TITLE OF INVENTION: No. 5990383el Pesticidal Proteins and Strains
: NUMBER OF SEQUENCES: 50
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CIBA-GEIGY Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
:
COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentn Release #1.0, Version #1.30B
:
CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/469,334
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 800
:
PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/463,483
: FILING DATE:
: APPLICATION NUMBER: US 08/314,594
: FILING DATE: 09-SEP-1994
:
PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/218,018
: FILING DATE: 23-MAR-1994
:
PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/037,057
: FILING DATE: 25-MAR-1993
:
ATTORNEY/AGENT INFORMATION:
: NAME: Sprull, W. Murray
: REGISTRATION NUMBER: 32,943
: REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8615
: TELEFAX: 919-541-8689
:
INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 884 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
:
MOLECULE TYPE: protein
:
US-08-469-334-5

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:49:24 ; Search time 55.64 Seconds
(without alignments)
110.818 Million cell updates/sec

Title: US-09-747-521-2_COPY_9_282
Perfect score: 1404
Sequence: 1 KVISMSCLVATITLSGPVFI.....RDVLQLYAPEAFNYMDKENE 274

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1282	91.3	456	1	US-08-021-601-6
2	1282	91.3	456	1	US-08-082-849B-6
3	1282	91.3	456	5	PCT-US94-01624-6
4	1282	91.3	472	1	US-08-021-601-8
5	1282	91.3	472	1	US-08-082-849B-8
6	1282	91.3	472	5	PCT-US94-01624-8
7	1282	91.3	508	1	US-08-021-601-10
8	1282	91.3	508	1	US-08-082-849B-10
9	1282	91.3	508	5	PCT-US94-01624-10
10	1282	91.3	776	1	US-08-021-601-2
11	1282	91.3	776	1	US-08-082-849B-2
12	1282	91.3	776	6	PCT-US94-01624-2
13	442	91.3	800	6	PCT-US94-01624-2
14	113	8.0	593	1	US-07-961-522-4
15	113	8.0	593	1	US-08-217-438-4
16	113	8.0	593	1	US-08-217-438-5
17	113	8.0	593	1	US-08-487-850A-100
18	113	8.0	593	1	US-08-321-978-4
19	113	8.0	593	2	US-08-710-584-4
20	113	8.0	593	2	US-08-478-435-100
21	113	8.0	593	2	US-08-337-483-100
22	113	8.0	593	2	US-08-478-373-100
23	113	8.0	593	3	US-08-474-671-100
24	113	8.0	593	3	US-08-483-577A-100
25	113	8.0	593	4	US-08-897-438-100
26	112	8.0	1786	4	US-08-973-462-8
27	111	7.9	700	1	US-07-720-589-2

28	111	7.9	700	2	US-08-785-190-2	Sequence 2, Appli
29	111	7.9	700	4	US-08-235-836C-66	Sequence 66, Appli
30	111	7.9	700	5	PCT-US92-05539-2	Sequence 2, Appli
31	111	7.9	1104	4	US-08-923-992A-4	Sequence 4, Appli
32	111	7.9	1164	4	US-08-923-992A-2	Sequence 2, Appli
33	108.5	7.7	937	1	US-08-253-155A-31	Sequence 31, Appli
34	108.5	7.7	1164	4	US-08-923-992A-10	Sequence 10, Appli
35	107	7.6	3111	2	US-08-460-309-4	Sequence 4, Appli
36	107	7.6	3111	2	US-08-125-077-4	Sequence 4, Appli
37	106.5	7.6	381	2	US-08-858-052-3	Sequence 3, Appli
38	106.5	7.6	381	3	US-09-200-284-3	Sequence 3, Appli
39	106.5	7.6	1588	5	PCT-US93-07261-11	Sequence 11, Appli
40	106.5	7.6	1663	5	PCT-US93-07261-16	Sequence 16, Appli
41	106	7.5	976	4	US-09-104-324B-4	Sequence 4, Appli
42	106	7.5	1098	4	US-08-923-992A-8	Sequence 8, Appli
43	105.5	7.5	732	2	US-08-533-669A-18	Sequence 18, Appli
44	102.5	7.3	1128	4	US-08-923-992A-6	Sequence 6, Appli
45	102	7.3	337	1	US-08-445-135-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-021-601-6
Sequence 6, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Atoria, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-9880
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-6

Query Match 91.3%; Score 1282; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 6.6e-98;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGHGIVGHNHVEKKNKRNKRKDEIRKJTOEHLKELMKHIVKTEVAGEAVKKEAE 85

Db 1 AGGHGDMHVKKEKKDKDKRDEERNKTOEHLKEIKKHIVKIEVKGEAVKKEAAE 60
QY 86 KLEKVPDVLKEMKKAIGKIYIVDGDITKHISLEALSEDKKKIKIDYGDALLHEHYV 145
Db 61 KLEKVPDVLKEMKKAIGKIYIVDGDITKHISLEALSEDKKKIKIDYGDALLHEHYV 120
QY 146 AKEGEPLVLIQSSSEDEVVENTERKALNYYEIGKILSRDILSKINQPYOKFLDVLNTIKNA 205
Db 121 AKEGEPLVLIQSSSEDEVVENTERKALNYYEIGKILSRDILSKINQPYOKFLDVLNTIKNA 180
QY 206 SDSGODLLFTNOLKEHPTDPSVEFLQNSNEVOEYAKAFAYIIEPOHHDVLIQVAPEA 265
Db 181 SDSGODLLFTNOLKEHPTDPSVEFLQNSNEVOEYAKAFAYIIEPOHHDVLIQVAPEA 240
QY 266 FNYMDKENE 274
Db 241 FNYMDKENE 249

RESULT 2

US-08-082-849B-6
Sequence 6, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Ariora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
TITLE OF INVENTION: Related Methods
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-6

Query Match 91.3%; Score 1282; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 6, 6e-98;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 AGGHGDMHVKKEKKDKDKRDEERNKTOEHLKEIKKHIVKIEVKGEAVKKEAAE 85

Db 1 AGGHGDMHVKKEKKDKDKRDEERNKTOEHLKEIKKHIVKIEVKGEAVKKEAAE 60
QY 86 KLEKVPDVLKEMKKAIGKIYIVDGDITKHISLEALSEDKKKIKIDYGDALLHEHYV 145
Db 61 KLEKVPDVLKEMKKAIGKIYIVDGDITKHISLEALSEDKKKIKIDYGDALLHEHYV 120
QY 146 AKEGEPLVLIQSSSEDEVVENTERKALNYYEIGKILSRDILSKINQPYOKFLDVLNTIKNA 205
Db 121 AKEGEPLVLIQSSSEDEVVENTERKALNYYEIGKILSRDILSKINQPYOKFLDVLNTIKNA 180
QY 206 SDSGODLLFTNOLKEHPTDPSVEFLQNSNEVOEYAKAFAYIIEPOHHDVLIQVAPEA 265
Db 181 SDSGODLLFTNOLKEHPTDPSVEFLQNSNEVOEYAKAFAYIIEPOHHDVLIQVAPEA 240
QY 266 FNYMDKENE 274
Db 241 FNYMDKENE 249

RESULT 3

PCT-US94-01624-6
Sequence 6, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Ariora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESSES:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: Stewart Street Tower, 20th Floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-6

Query Match 91.3%; Score 1282; DB 5; Length 456;
Best Local Similarity 100.0%; Pred. No. 6, 6e-98;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 AGGHGDMHVKKEKKDKDKRDEERNKTOEHLKEIKKHIVKIEVKGEAVKKEAAE 85
Db 1 AGGHGDMHVKKEKKDKDKRDEERNKTOEHLKEIKKHIVKIEVKGEAVKKEAAE 60

OY 86 KLEKVPSPVLEMYKAIGKIIYIVDGDITKHSLSLEALSDKKKKIKDIYKDALLHEHYV 145
|
DB 61 KLEKVPSPVLEMYKAIGKIIYIVDGDITKHSLSLEALSDKKKKIKDIYKDALLHEHYV 120
OY 146 AKEGEPVLVIQSSSEDIYVENTEKALNVYYEIGKILSRDLISKINPYQKFLDVLNTIKNA 205
|
DB 121 AKEGEPVLVIQSSSEDIYVENTEKALNVYYEIGKILSRDLISKINPYQKFLDVLNTIKNA 180
OY 206 SSDSDODLFTNQLKHPDPSVEFLEQNSNEVOEYFAKAFAYIIEPQRDVLTQYAPPA 265
|
DB 181 SSDSDODLFTNQLKHPDPSVEFLEQNSNEVOEYFAKAFAYIIEPQRDVLTQYAPPA 240
OY 266 FNYMDKFNE 274
|
DB 241 FNYMDKFNE 249

RESULT 4
US-08-021-601-8
; Sequence 8, Application US/08021601
; Patent No. 5591631
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Nichols, Peter J.
; APPLICANT: Atora, Naveen
; APPLICANT: Singh, Yogendra
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021.601
; FILING DATE: 19930212
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-021-601-8

Query Match 91.3%; Score 1282; DB 1; Length 472;
Best Local Similarity 100.0%; Pred. No. 6.9e-98;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 26 AGHGQVGNHNVKEKKNKDEBNKKTQDEHLKEIKHIVKIEVGEAEVKKKEAE 85
|
DB 4 AGHGQVGNHNVKEKKNKDEBNKKTQDEHLKEIKHIVKIEVGEAEVKKKEAE 63
OY 86 KLEKVPSPVLEMYKAIGKIIYIVDGDITKHSLSLEALSDKKKKIKDIYKDALLHEHYV 145
|
DB 64 KLEKVPSPVLEMYKAIGKIIYIVDGDITKHSLSLEALSDKKKKIKDIYKDALLHEHYV 123

OY 146 AKEGEPVLVIQSSSEDIYVENTEKALNVYYEIGKILSRDLISKINPYQKFLDVLNTIKNA 205
|
DB 124 AKEGEPVLVIQSSSEDIYVENTEKALNVYYEIGKILSRDLISKINPYQKFLDVLNTIKNA 183
OY 206 SSDSDODLFTNQLKHPDPSVEFLEQNSNEVOEYFAKAFAYIIEPQRDVLTQYAPPA 265
|
DB 184 SSDSDODLFTNQLKHPDPSVEFLEQNSNEVOEYFAKAFAYIIEPQRDVLTQYAPPA 243
OY 266 FNYMDKFNE 274
|
DB 244 FNYMDKFNE 252

RESULT 5
US-08-082-849B-8
; Sequence 8, Application US/08082849B
; Patent No. 5677274
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Atora, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/082.849B
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/021,601
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-161-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-082-849B-8

Query Match 91.3%; Score 1282; DB 1; Length 472;
Best Local Similarity 100.0%; Pred. No. 6.9e-98;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 26 AGHGQVGNHNVKEKKNKDEBNKKTQDEHLKEIKHIVKIEVGEAEVKKKEAE 85
|
DB 4 AGHGQVGNHNVKEKKNKDEBNKKTQDEHLKEIKHIVKIEVGEAEVKKKEAE 63
OY 86 KLEKVPSPVLEMYKAIGKIIYIVDGDITKHSLSLEALSDKKKKIKDIYKDALLHEHYV 145
|
DB 64 KLEKVPSPVLEMYKAIGKIIYIVDGDITKHSLSLEALSDKKKKIKDIYKDALLHEHYV 123

QY 146 AKEGEYEVLYIOSEEDYVENTEKALNYYEIGKILSRDIISKINOPYOKFLDVLNTIKNA 205
DB 124 AKEGEYEVLYIOSEEDYVENTEKALNYYEIGKILSRDIISKINOPYOKFLDVLNTIKNA 183
QY 206 SDSGQDLFTNOLKEHPDTSVEFLSEONSNEVOEFAKAFAYIIEPOHRDVLDTYAPEA 265
DB 184 SDSGQDLFTNOLKEHPDTSVEFLSEONSNEVOEFAKAFAYIIEPOHRDVLDTYAPEA 243
QY 266 FNYMDKENE 274
DB 244 FNYMDKENE 252

RESULT 6
PCT-US94-01624-8
Sequence 8, Application PC/TUS9401624

GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
RELATED METHODS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESSES:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: Stewart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-8

Query Match 91.3%; Score 1282; DB 5; Length 472;
Best Local Similarity 100.0%; Pred. No. 6.9e-98;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGHGVDGMHYKKEKKNDKRRKDEERNKTOEHLKEIMKHIVKIEVKGGEAVKKEAAE 85
DB 4 AGHGVDGMHYKKEKKNDKRRKDEERNKTOEHLKEIMKHIVKIEVKGGEAVKKEAAE 63
QY 86 KLEKVPDVLKEMKATGKIYIVDGTITKHSLEALSEDKKTKTDYGDALLHEHYV 145
DB 64 KLEKVPDVLKEMKATGKIYIVDGTITKHSLEALSEDKKTKTDYGDALLHEHYV 123
QY 146 AKEGEYEVLYIOSEEDYVENTEKALNYYEIGKILSRDIISKINOPYOKFLDVLNTIKNA 205

DB 124 AKEGEYEVLYIOSEEDYVENTEKALNYYEIGKILSRDIISKINOPYOKFLDVLNTIKNA 183
QY 206 SDSGQDLFTNOLKEHPDTSVEFLSEONSNEVOEFAKAFAYIIEPOHRDVLDTYAPEA 265
DB 184 SDSGQDLFTNOLKEHPDTSVEFLSEONSNEVOEFAKAFAYIIEPOHRDVLDTYAPEA 243
QY 266 FNYMDKENE 274
DB 244 FNYMDKENE 252

RESULT 7
US-08-021-601-10
Sequence 10, Application US/08021601
Patent No. 5591631

GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 508 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-10

Query Match 91.3%; Score 1282; DB 1; Length 508;
Best Local Similarity 100.0%; Pred. No. 7.5e-98;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGHGVDGMHYKKEKKNDKRRKDEERNKTOEHLKEIMKHIVKIEVKGGEAVKKEAAE 85
DB 1 AGHGVDGMHYKKEKKNDKRRKDEERNKTOEHLKEIMKHIVKIEVKGGEAVKKEAAE 60
QY 86 KLEKVPDVLKEMKATGKIYIVDGTITKHSLEALSEDKKTKTDYGDALLHEHYV 145
DB 61 KLEKVPDVLKEMKATGKIYIVDGTITKHSLEALSEDKKTKTDYGDALLHEHYV 120
QY 146 AKEGEYEVLYIOSEEDYVENTEKALNYYEIGKILSRDIISKINOPYOKFLDVLNTIKNA 205
DB 121 AKEGEYEVLYIOSEEDYVENTEKALNYYEIGKILSRDIISKINOPYOKFLDVLNTIKNA 180
QY 206 SDSGQDLFTNOLKEHPDTSVEFLSEONSNEVOEFAKAFAYIIEPOHRDVLDTYAPEA 265

DB 181 SPSDGDLLFTNQLKHPDTSVEFLQNSNEVQEFVFAKAFAYIIPQHRDVLQLYAPEA 240
QY 266 FNYMDKFNE 274
DB 241 FNYMDKFNE 249

RESULT 8

US-08-082-849B-10
; Sequence 10, Application US/08082849B
; Patent No. 5677274
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Atora, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
; TITLE OF INVENTION: Related Methods
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/082, 849B
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/021,601
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-161-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 508 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-082-849B-10

Query Match 91.3%; Score 1282; DB 1; Length 508;
Best Local Similarity 100.0%; Pred. No. 7.5e-98;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGCHDVGAMHVEKEKKNKDEERNKKTQBEHLKEIMKHIVKLEVGEEVAKKEAAE 85
DB 1 AGCHDVGAMHVEKEKKNKDEERNKKTQBEHLKEIMKHIVKLEVGEEVAKKEAAE 60
QY 86 KLEKVPDYLEMYKAIGKIYIVDGDITKHISLEALSEDKKIKDIYKDALHHEHYV 145
DB 61 KLEKVPDYLEMYKAIGKIYIVDGDITKHISLEALSEDKKIKDIYKDALHHEHYV 120
QY 146 AKEGYEPVAVIOSSEEDYVENTEKALNVYIEIGKILSRDILSKINQYOKFLDVLNTIKNA 205
DB 121 AKEGYEPVAVIOSSEEDYVENTEKALNVYIEIGKILSRDILSKINQYOKFLDVLNTIKNA 180
QY 206 SPSDGDLLFTNQLKHPDTSVEFLQNSNEVQEFVFAKAFAYIIPQHRDVLQLYAPEA 265

DB 181 SPSDGDLLFTNQLKHPDTSVEFLQNSNEVQEFVFAKAFAYIIPQHRDVLQLYAPEA 240
QY 266 FNYMDKFNE 274
DB 241 FNYMDKFNE 249

RESULT 9

PCT-US94-01624-10
; Sequence 10, Application PC/TUS9401624
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Atora, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
; STREET: Stewart Street Tower, 20th Floor, One Market
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01624
; FILING DATE: June 25, 1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-115
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 508 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-01624-10

Query Match 91.3%; Score 1282; DB 5; Length 508;
Best Local Similarity 100.0%; Pred. No. 7.5e-98;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGCHDVGAMHVEKEKKNKDEERNKKTQBEHLKEIMKHIVKLEVGEEVAKKEAAE 85
DB 1 AGCHDVGAMHVEKEKKNKDEERNKKTQBEHLKEIMKHIVKLEVGEEVAKKEAAE 60
QY 86 KLEKVPDYLEMYKAIGKIYIVDGDITKHISLEALSEDKKIKDIYKDALHHEHYV 145
DB 61 KLEKVPDYLEMYKAIGKIYIVDGDITKHISLEALSEDKKIKDIYKDALHHEHYV 120
QY 146 AKEGYEPVAVIOSSEEDYVENTEKALNVYIEIGKILSRDILSKINQYOKFLDVLNTIKNA 205
DB 121 AKEGYEPVAVIOSSEEDYVENTEKALNVYIEIGKILSRDILSKINQYOKFLDVLNTIKNA 180
QY 206 SPSDGDLLFTNQLKHPDTSVEFLQNSNEVQEFVFAKAFAYIIPQHRDVLQLYAPEA 265
DB 181 SPSDGDLLFTNQLKHPDTSVEFLQNSNEVQEFVFAKAFAYIIPQHRDVLQLYAPEA 240

QY 266 FNYMDKFN 274
|
Db 241 FNYMDKFN 249

RESULT 10
US-08-021-601-2

; Sequence 2, Application US/08021601
; Patent No. 5591631
; GENERAL INFORMATION:
; APPLICANT: Leppia, Stephen H.
; APPLICANT: Kilmpel, Kurt R.
; APPLICANT: Nichols, Peter J.
; APPLICANT: Aroora, Naveen
; APPLICANT: Singh, Yogendra
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,601
; FILING DATE: 19930212
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 776 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-021-601-2

Query Match 91.3%; Score 1282; DB 1; Length 776;
Best Local Similarity 100.0%; Pred. No. 1.3e-97;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 AGHGDVGMVKKKKKKKKRDEERNKTOEHLKEIMKHIVKIVKKEEAVKKAEE 85
|
Db 1 AGHGDVGMVKKKKKKKKRDEERNKTOEHLKEIMKHIVKIVKKEEAVKKAEE 60
QY 86 KLEKVPDVLKEMVKAIGKIYVDGDTKHSISLEALSEDKKKIKDYGKDALHHEHYV 145
|
Db 61 KLEKVPDVLKEMVKAIGKIYVDGDTKHSISLEALSEDKKKIKDYGKDALHHEHYV 120
QY 146 AKEGYEVLVIOSSSEDEVVENTERKALNYYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 205
|
Db 121 AKEGYEVLVIOSSSEDEVVENTERKALNYYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 180
QY 206 SDSGDQLLFTTNQKKEHPDPSVEFLQNSNEVOEPAKAFAYITIEQHRDVLQIAPEA 265
|
Db 181 SDSGDQLLFTTNQKKEHPDPSVEFLQNSNEVOEPAKAFAYITIEQHRDVLQIAPEA 240
QY 266 FNYMDKFN 274
|
Db 241 FNYMDKFN 249

RESULT 11

US-08-082-849B-2
; Sequence 2, Application US/08082849B
; Patent No. 5677274
; GENERAL INFORMATION:
; APPLICANT: Leppia, Stephen H.
; APPLICANT: Kilmpel, Kurt R.
; APPLICANT: Aroora, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/082,849B
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/021,601
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-161-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 776 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-082-849B-2

Query Match 91.3%; Score 1282; DB 1; Length 776;
Best Local Similarity 100.0%; Pred. No. 1.3e-97;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 AGHGDVGMVKKKKKKKKRDEERNKTOEHLKEIMKHIVKIVKKEEAVKKAEE 85
|
Db 1 AGHGDVGMVKKKKKKKKRDEERNKTOEHLKEIMKHIVKIVKKEEAVKKAEE 60
QY 86 KLEKVPDVLKEMVKAIGKIYVDGDTKHSISLEALSEDKKKIKDYGKDALHHEHYV 145
|
Db 61 KLEKVPDVLKEMVKAIGKIYVDGDTKHSISLEALSEDKKKIKDYGKDALHHEHYV 120
QY 146 AKEGYEVLVIOSSSEDEVVENTERKALNYYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 205
|
Db 121 AKEGYEVLVIOSSSEDEVVENTERKALNYYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 180
QY 206 SDSGDQLLFTTNQKKEHPDPSVEFLQNSNEVOEPAKAFAYITIEQHRDVLQIAPEA 265
|
Db 181 SDSGDQLLFTTNQKKEHPDPSVEFLQNSNEVOEPAKAFAYITIEQHRDVLQIAPEA 240
QY 266 FNYMDKFN 274
|
Db 241 FNYMDKFN 249

TELEPHONE: (415) 617-8999
TELEFAX: (415) 327-3231
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 593 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-961-522-4

Query Match 8.0%; Score 113; DB 1; Length 593;
Best Local Similarity 24.3%; Pred. No. 0.14; Mismatches 91; Indels 88; Gaps 19;
Matches 72; Conservative 45;

22 LVQAGGHDGVGMH-VKEEKNKDE---NKRKDEERNKQOEHLKEIMKHIVKIE---VK 74
17 LVACSGGSGFDELDVAPNQTAKAEKATTSYODEETKKTKEELDKLMEPALGYEQILR 76
75 GEEAVKKEAEKLEKVPSPVLEMYKAIGKIYIVDGDITKHISLEALSDKKKINDIYG 134
77 RNKAPTEETEKKNER---VVELSEDKITKLYOESVEIIPH--LDELN-GKTTSDVY- 128
135 KDALLHEH-----YYAKEGEPEVLVIQSEEDYVENTEKALN-----YYY- 174
129 ---SHDSKRLDKNRDLKYVRSGY---VYDGSFNEIRNDSGFHFVKOGIDGYVYL 179
175 -----EI--GKILSR---DILSKINQPYQKFLVNLTI---KNAS-----DSD 209
180 GVTPEKELPKGKVIYSKGTWDFVSNIN--LEREIDGFDTSGDCKNVSATSITETVNDHK 237
210 GDDLLFTNOLK--EHPDSEVFLEQ-----NSNEVOEFAKAFAYIE 251
238 VGEKLDNEVKGVAHSEFAVDNKKLTGSLYRNGYINRNKAOEYTKR---YSIE 290

RESULT 15

US-08-217-438-4
Sequence 4, Application US/08217438
Patent No. 5521072

GENERAL INFORMATION:
APPLICANT: Potter, Andrew A.
APPLICANT: Gerlach, Gerald F.
APPLICANT: Willson, Philip J.
APPLICANT: Rossi-Campos, Amalia
TITLE OF INVENTION: ACTINOBACILLUS PLEUROPEIMONIAE
TITLE OF INVENTION: TRANSFERIN BINDING PROTEINS AND USES THEREOF
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Reed & Robins
STREET: 635 Bryant Street
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94301

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/217,438
FILING DATE: 22-MAR-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Robins, Roberta L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9001-0015.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-617-8999
TELEFAX: 415-327-3231
TELEX:

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 593 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-217-438-4

Query Match 8.0%; Score 113; DB 1; Length 593;
Best Local Similarity 24.3%; Pred. No. 0.14; Mismatches 91; Indels 88; Gaps 19;
Matches 72; Conservative 45;

22 LVQAGGHDGVGMH-VKEEKNKDE---NKRKDEERNKQOEHLKEIMKHIVKIE---VK 74
17 LVACSGGSGFDELDVAPNQTAKAEKATTSYODEETKKTKEELDKLMEPALGYEQILR 76
75 GEEAVKKEAEKLEKVPSPVLEMYKAIGKIYIVDGDITKHISLEALSDKKKINDIYG 134
77 RNKAPTEETEKKNER---VVELSEDKITKLYOESVEIIPH--LDELN-GKTTSDVY- 128
135 KDALLHEH-----YYAKEGEPEVLVIQSEEDYVENTEKALN-----YYY- 174
129 ---SHDSKRLDKNRDLKYVRSGY---VYDGSFNEIRNDSGFHFVKOGIDGYVYL 179
175 -----EI--GKILSR---DILSKINQPYQKFLVNLTI---KNAS-----DSD 209
180 GVTPEKELPKGKVIYSKGTWDFVSNIN--LEREIDGFDTSGDCKNVSATSITETVNDHK 237
210 GDDLLFTNOLK--EHPDSEVFLEQ-----NSNEVOEFAKAFAYIE 251
238 VGEKLDNEVKGVAHSEFAVDNKKLTGSLYRNGYINRNKAOEYTKR---YSIE 290

Search completed: December 2, 2001, 13:49:25
Job time: 205 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:49:43 ; Search time 55.64 Seconds
(without alignments)
225.680 Million cell updates/sec

Title: US-09-747-521-4_COPY_178_735
Perfect score: 2871
Sequence: 1 DNQLPELKKKSSNSRRKRS.....LYISNPKNVAVTKENT 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/2/1aa/6C.COMB.pep:*
6: /cgn2_6/prodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	2758	96.1	735 1 US-08-021-601-4	Sequence 4, Appl1
2	2758	96.1	735 1 US-08-082-849B-4	Sequence 4, Appl1
3	2758	96.1	735 5 PCT-US94-01624-4	Sequence 4, Appl1
4	2758	96.1	903 1 US-08-021-601-12	Sequence 12, Appl1
5	2758	96.1	903 1 US-08-082-849B-12	Sequence 12, Appl1
6	2758	96.1	903 5 PCT-US94-01624-12	Sequence 12, Appl1
7	2581	89.9	719 1 US-08-082-849B-31	Sequence 31, Appl1
8	2581	89.9	719 5 PCT-US94-01624-31	Sequence 31, Appl1
9	591	20.6	881 4 US-08-960-780-32	Sequence 32, Appl1
10	591	20.6	881 4 US-09-073-898-32	Sequence 32, Appl1
11	589	20.5	884 1 US-08-471-033-5	Sequence 5, Appl1
12	589	20.5	884 2 US-08-471-044-5	Sequence 5, Appl1
13	589	20.5	884 2 US-08-471-044-5	Sequence 5, Appl1
14	589	20.5	884 2 US-08-471-046A-5	Sequence 5, Appl1
15	589	20.5	884 2 US-08-470-566B-5	Sequence 5, Appl1
16	589	20.5	884 2 US-08-469-334-5	Sequence 5, Appl1
17	589	20.5	884 3 US-09-300-529-5	Sequence 5, Appl1
18	589	20.5	1346 1 US-08-471-033-23	Sequence 23, Appl1
19	589	20.5	1346 2 US-08-471-044-23	Sequence 23, Appl1
20	589	20.5	1346 2 US-08-463-483A-23	Sequence 23, Appl1
21	589	20.5	1346 2 US-08-471-046A-23	Sequence 23, Appl1
22	589	20.5	1346 2 US-08-470-566B-23	Sequence 23, Appl1
23	589	20.5	1346 2 US-08-469-334-23	Sequence 23, Appl1
24	589	20.5	1346 3 US-09-300-529-23	Sequence 23, Appl1
25	588	20.5	852 1 US-08-471-033-36	Sequence 36, Appl1
26	588	20.5	852 2 US-08-471-044-36	Sequence 36, Appl1
27	588	20.5	852 2 US-08-463-483A-36	Sequence 36, Appl1

28	588	20.5	852 2 US-08-471-046A-36	Sequence 36, Appl1
29	588	20.5	852 2 US-08-470-566B-36	Sequence 36, Appl1
30	588	20.5	852 2 US-08-469-334-36	Sequence 36, Appl1
31	588	20.5	852 3 US-09-300-529-36	Sequence 36, Appl1
32	588	20.5	1338 1 US-08-471-033-50	Sequence 50, Appl1
33	588	20.5	1338 2 US-08-471-044-50	Sequence 50, Appl1
34	588	20.5	1338 2 US-08-463-483A-50	Sequence 50, Appl1
35	588	20.5	1338 2 US-08-471-046A-50	Sequence 50, Appl1
36	588	20.5	1338 2 US-08-470-566B-50	Sequence 50, Appl1
37	588	20.5	1338 2 US-08-469-334-50	Sequence 50, Appl1
38	588	20.5	1338 3 US-09-300-529-50	Sequence 50, Appl1
39	582.5	20.3	784 4 US-09-371-913A-7	Sequence 7, Appl1
40	580	20.2	667 1 US-08-471-033-7	Sequence 7, Appl1
41	580	20.2	667 2 US-08-471-044-7	Sequence 7, Appl1
42	580	20.2	667 2 US-08-463-483A-7	Sequence 7, Appl1
43	580	20.2	667 2 US-08-471-046A-7	Sequence 7, Appl1
44	580	20.2	667 2 US-08-470-566B-7	Sequence 7, Appl1
45	580	20.2	667 2 US-08-469-334-7	Sequence 7, Appl1

ALIGNMENTS

RESULT 1
US-08-021-601-4
Sequence 4, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppia, Stephen H.
APPLICANT: Kilmpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Anora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-4

Query Match 96.1%; Score 2758; DB 1; Length 735;
Best Local Similarity 96.8%; Pred. No. 3.7e-213;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;
QY 1 DNQLPELKKKSSNSRRKRSACPTVPDDNDGICPDSLEVEGTYVKKKRTFLSPWIS 60

```
Db 149 DNLQJPELKQKSSNSRRKRSSTAGPTVPDRDNDGIPDSLEVEGYTVYDVKNKRTFLSPWIS 208
Qy 61 NIHEKKGLTKYKSSPEKSTASDPYSDPEKYVGRIDKNVSPARHPVLAAYPIVHVDMEN 120
Db 209 NIHEKKGLTKYKSSPEKSTASDPYSDPEKYVGRIDKNVSPARHPVLAAYPIVHVDMEN 268
Qy 121 ILSKNEOSTONTSEFTTISKNTSTRTHTSEVHGNAEVANSTSTRTHTSEVHGNAE 180
Db 269 ILSKNEOSTONTSEFTTISKNTSTRTHTSEVHGNAEVANSTSTRTHTSEVHGNAE 328
Qy 181 VHAVIDHSLSLAGERETAETGLMTADTARLANIRVYNTGTAPLYNVLPTTSLVLGN 240
Db 329 SSTVAIDHSLSLAGERETAETGLMTADTARLANIRVYNTGTAPLYNVLPTTSLVLGN 388
Qy 241 QTLATIKAKENQSOILAPNNYPSKNIAPALNADDFSSPTITMNNQFLEKTKQL 300
Db 389 QTLATIKAKENQSOILAPNNYPSKNIAPALNADDFSSPTITMNNQFLEKTKQL 448
Qy 301 RLDTQVYGNATYVNFENGRAVDGSMNSEVLPQIOETTARIIFNGKDLNVERRIA 360
Db 449 RLDTQVYGNATYVNFENGRAVDGSMNSEVLPQIOETTARIIFNGKDLNVERRIA 508
Qy 361 NPSDPLETTKPDMTLKEALKIAFGNEPNGNLQYOGKDTTEPFNFDOOTSONIKNO 420
Db 509 NPSDPLETTKPDMTLKEALKIAFGNEPNGNLQYOGKDTTEPFNFDOOTSONIKNO 568
Qy 421 LNATNITYVLDKIKILNAKNNILIRDKRFHYDRNNIYAVGADSEVVAEAREVINS 480
Db 569 LNATNITYVLDKIKILNAKNNILIRDKRFHYDRNNIYAVGADSEVVAEAREVINS 628
Qy 481 LNIDDKIRKILSGYIVETEDTEGLKEVINDRYDMLNSSLRODGKTFIDFKKYN 540
Db 629 LNIDDKIRKILSGYIVETEDTEGLKEVINDRYDMLNSSLRODGKTFIDFKKYN 688
Qy 541 ISNPYKVVAVYATKENT 558
Db 689 ISNPYKVVAVYATKENT 706

RESULT 2
US-08-082-849B-4
: Sequence 4, Application US/08082849B
: Patent No. 5677274
: GENERAL INFORMATION:
: APPLICANT: Leppla, Stephen H.
: APPLICANT: Kimpel, Kurt R.
: APPLICANT: Arota, Naveen
: APPLICANT: Singh, Yogendra
: APPLICANT: Nichols, Peter J.
: TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
: NUMBER OF SEQUENCES: 35
: CORRESPONDENCE ADDRESS:
: ADDRESS: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/082,849B
: FILING DATE: 25-JUN-1993
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/021,601
: FILING DATE: 12-FEB-1993
: ATTORNEY/AGENT INFORMATION:
```

```
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-4

Query Match          96.1%; Score 2758; DB 1: Length 735;
Best Local Similarity 96.8%; Pred. No. 3,7e+213;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

Db 1 DNLQJPELKQKSSNSRRKRSSTAGPTVPDRDNDGIPDSLEVEGYTVYDVKNKRTFLSPWIS 60
Db 149 DNLQJPELKQKSSNSRRKRSSTAGPTVPDRDNDGIPDSLEVEGYTVYDVKNKRTFLSPWIS 208
Qy 61 NIHEKKGLTKYKSSPEKSTASDPYSDPEKYVGRIDKNVSPARHPVLAAYPIVHVDMEN 120
Db 209 NIHEKKGLTKYKSSPEKSTASDPYSDPEKYVGRIDKNVSPARHPVLAAYPIVHVDMEN 268
Qy 121 ILSKNEOSTONTSEFTTISKNTSTRTHTSEVHGNAEVANSTSTRTHTSEVHGNAE 180
Db 269 ILSKNEOSTONTSEFTTISKNTSTRTHTSEVHGNAEVANSTSTRTHTSEVHGNAE 328
Qy 181 VHAVIDHSLSLAGERETAETGLMTADTARLANIRVYNTGTAPLYNVLPTTSLVLGN 240
Db 329 SSTVAIDHSLSLAGERETAETGLMTADTARLANIRVYNTGTAPLYNVLPTTSLVLGN 388
Qy 241 QTLATIKAKENQSOILAPNNYPSKNIAPALNADDFSSPTITMNNQFLEKTKQL 300
Db 389 QTLATIKAKENQSOILAPNNYPSKNIAPALNADDFSSPTITMNNQFLEKTKQL 448
Qy 481 LNIDDKIRKILSGYIVETEDTEGLKEVINDRYDMLNSSLRODGKTFIDFKKYN 540
Db 629 LNIDDKIRKILSGYIVETEDTEGLKEVINDRYDMLNSSLRODGKTFIDFKKYN 688
Qy 541 ISNPYKVVAVYATKENT 558
Db 689 ISNPYKVVAVYATKENT 706

RESULT 3
PCT-US94-01624-4
: Sequence 4, Application PC/TUS9401624
: GENERAL INFORMATION:
: APPLICANT: Leppla, Stephen H.
: APPLICANT: Kimpel, Kurt R.
: APPLICANT: Arota, Naveen
: APPLICANT: Singh, Yogendra
: APPLICANT: Nichols, Peter J.
: TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
```

ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
STREET: Steuart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCF-US94-01624-4

Query Match 96.1%; Score 2758; DB 5; Length 735;
Best Local Similarity 96.8%; Pred. No. 3.7e-213;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

1 DNQLPELKOKSSNRKRSTASGAPVPPDRDNDGIPDSLEVEGYTVVKKRRTFLSPWIS 60
149 DNQLPELKOKSSNRKRSTASGAPVPPDRDNDGIPDSLEVEGYTVVKKRRTFLSPWIS 208
61 NHEKKGLTKYKSSPEKWSASDPYSDFEKYTGRIKKNVSPEARHPLVAAPIYHVDMEN 120
209 NHEKKGLTKYKSSPEKWSASDPYSDFEKYTGRIKKNVSPEARHPLVAAPIYHVDMEN 268
121 IILSKNEDOSTQNTDSETRTISKNTSRTHTSEVHGNAEYHANTSTSRHTSEVHGNAE 180
269 IILSKNEDOSTQNTDSETRTISKNTSRTHTSEVHGNAEYHANTSTSRHTSEVHGNAE 328
181 VHAVAIDHSLSLAGERTWAEETMGLNTADTARLANIRYVNTGTAPRYNVLPPTSLSVLGKN 240
329 SSTVAIDHSLSLAGERTWAEETMGLNTADTARLANIRYVNTGTAPRYNVLPPTSLSVLGKN 388
241 QTLATIRAKENQSLIAPNNYPSKNLAPIALNAODFSSPTITMNYNOFLELEKTQOL 300
389 QTLATIRAKENQSLIAPNNYPSKNLAPIALNAODFSSPTITMNYNOFLELEKTQOL 448
301 RLDTDOYVGNIAIYFENGVRVDTGSMNSEVLPQIDETTARILFNKDLNLVERRIAAV 360
449 RLDTDOYVGNIAIYFENGVRVDTGSMNSEVLPQIDETTARILFNKDLNLVERRIAAV 508
361 NPSPLETTKPDMLTKALIAFGFNPENGNLOYGKDITEFDENFPFOOTSONTKKNOLAE 420
509 NPSPLETTKPDMLTKALIAFGFNPENGNLOYGKDITEFDENFPFOOTSONTKKNOLAE 568
421 LNAFTNIYVLDKIRLAKNAILIRDKRFHYDRNNIAVGADESYYKAEAREVINSSTGLL 480
569 LNAFTNIYVLDKIRLAKNAILIRDKRFHYDRNNIAVGADESYYKAEAREVINSSTGLL 628
481 LNDKDKRRLTSGYIVIEPTGELKEYINDRYDMLNTISSLRQDCKTIDFKKYNDKPLPY 540
629 LNDKDKRRLTSGYIVIEPTGELKEYINDRYDMLNTISSLRQDCKTIDFKKYNDKPLPY 688
541 ISNPNNYKVNNYAVATKENT 558

Db 689 ISNPNNYKVNNYAVATKENT 706

RESULT 4
US-08-021-601-12
Sequence 12, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppia, Stephen H.
APPLICANT: Kilmpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Arcora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratl, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-12

Query Match 96.1%; Score 2758; DB 1; Length 903;
Best Local Similarity 96.8%; Pred. No. 5e-213;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

1 DNQLPELKOKSSNRKRSTASGAPVPPDRDNDGIPDSLEVEGYTVVKKRRTFLSPWIS 60
149 DNQLPELKOKSSNRKRSTASGAPVPPDRDNDGIPDSLEVEGYTVVKKRRTFLSPWIS 208
61 NHEKKGLTKYKSSPEKWSASDPYSDFEKYTGRIKKNVSPEARHPLVAAPIYHVDMEN 120
209 NHEKKGLTKYKSSPEKWSASDPYSDFEKYTGRIKKNVSPEARHPLVAAPIYHVDMEN 268
121 IILSKNEDOSTQNTDSETRTISKNTSRTHTSEVHGNAEYHANTSTSRHTSEVHGNAE 180
269 IILSKNEDOSTQNTDSETRTISKNTSRTHTSEVHGNAEYHANTSTSRHTSEVHGNAE 328
181 VHAVAIDHSLSLAGERTWAEETMGLNTADTARLANIRYVNTGTAPRYNVLPPTSLSVLGKN 240
329 SSTVAIDHSLSLAGERTWAEETMGLNTADTARLANIRYVNTGTAPRYNVLPPTSLSVLGKN 388
241 QTLATIRAKENQSLIAPNNYPSKNLAPIALNAODFSSPTITMNYNOFLELEKTQOL 300
389 QTLATIRAKENQSLIAPNNYPSKNLAPIALNAODFSSPTITMNYNOFLELEKTQOL 448

QY 301 RLDTQVYGNATATYFENGRAVRVDGSMWSEVLPQIOETTARIIFNGKDLNVERIAAV 360
DB 449 RLDTQVYGNATATYFENGRAVRVDGSMWSEVLPQIOETTARIIFNGKDLNVERIAAV 508
QY 361 NPSDPLETTKPDMLKEALKIAFGNEPENGNIQYOGKDIETEDFENFDQOTSINIKNOAE 420
DB 509 NPSDPLETTKPDMLKEALKIAFGNEPENGNIQYOGKDIETEDFENFDQOTSINIKNOAE 568
QY 421 LNATNIYVLDKIKLNAMNIIIRDKRFHYDRNNIAGADESVKAEHREVINSSTEGIL 480
DB 569 LNATNIYVLDKIKLNAMNIIIRDKRFHYDRNNIAGADESVKAEHREVINSSTEGIL 628
QY 481 LNIDDKIRILSGYIIEIDTEGLEKVINDRYDMLNISLRDQGTFTIDFKKYNKDLPLY 540
DB 629 LNIDDKIRILSGYIIEIDTEGLEKVINDRYDMLNISLRDQGTFTIDFKKYNKDLPLY 688
QY 541 ISNPNKYVNVAVTKENT 558
DB 689 ISNPNKYVNVAVTKENT 706

RESULT 5

US-08-082-849B-12
Sequence 12, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Aroza, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-12

Query Match 96.1%; Score 2758; DB 1; Length 903;
Best Local Similarity 96.8%; Pred. No. 5e-213;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 1 DNLQPLKOKSSNRKRKRSAGPTVPDRDNDGIPDSLEVEGYVDYKKNKRTFLSPMJS 60
DB 149 DNLQPLKOKSSNRKRKRSAGPTVPDRDNDGIPDSLEVEGYVDYKKNKRTFLSPMJS 208
QY 61 NHEKKGLTKYKSSPEKKSTASDPYSDFEKYGRIDKNVSPARRPLVAAPIVHVDEN 120
DB 209 NHEKKGLTKYKSSPEKKSTASDPYSDFEKYGRIDKNVSPARRPLVAAPIVHVDEN 268
QY 121 ILSKNEQOSTONTDSETRTISKNTSTSRHTSEVHGAHVAHNTSRHTSEVHGAHAE 180
DB 269 ILSKNEQOSTONTDSETRTISKNTSTSRHTSEVHGAHVAHNTSRHTSEVHGAHAE 328
QY 161 VHAVALDHSLSIAGERTAEFTGLTADTARLANIRIVNTGTAPIYVLPPTSVLCKN 240
DB 329 SSTVAIDHSLSIAGERTAEFTGLTADTARLANIRIVNTGTAPIYVLPPTSVLCKN 388
QY 241 OTLATIKAKENQSOILAPNNYPSKNLAPIALNADDESSPTIMANTNOFLEKTKOL 300
DB 389 OTLATIKAKENQSOILAPNNYPSKNLAPIALNADDESSPTIMANTNOFLEKTKOL 448
QY 301 RLDTQVYGNATATYFENGRAVRVDGSMWSEVLPQIOETTARIIFNGKDLNVERIAAV 360
DB 449 RLDTQVYGNATATYFENGRAVRVDGSMWSEVLPQIOETTARIIFNGKDLNVERIAAV 508
QY 361 NPSDPLETTKPDMLKEALKIAFGNEPENGNIQYOGKDIETEDFENFDQOTSINIKNOAE 420
DB 509 NPSDPLETTKPDMLKEALKIAFGNEPENGNIQYOGKDIETEDFENFDQOTSINIKNOAE 568
QY 421 LNATNIYVLDKIKLNAMNIIIRDKRFHYDRNNIAGADESVKAEHREVINSSTEGIL 480
DB 569 LNATNIYVLDKIKLNAMNIIIRDKRFHYDRNNIAGADESVKAEHREVINSSTEGIL 628
QY 481 LNIDDKIRILSGYIIEIDTEGLEKVINDRYDMLNISLRDQGTFTIDFKKYNKDLPLY 540
DB 629 LNIDDKIRILSGYIIEIDTEGLEKVINDRYDMLNISLRDQGTFTIDFKKYNKDLPLY 688
QY 541 ISNPNKYVNVAVTKENT 558
DB 689 ISNPNKYVNVAVTKENT 706

RESULT 6

PCT-US94-01624-12
Sequence 12, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Aroza, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: Stewart Street Tower, 20th Floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677

REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-12

Query Match 96.1%; Score 2758; DB 5; Length 903;
Best Local Similarity 96.8%; Pred. No. 5e-213;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 1 DNLQPELKOKSSNSRRKSTASGPTVPDRDNDGIPDSLEVEGYVDVKKRFTLSFWIS 60
DB 149 DNLQPELKOKSSNSRRKSTASGPTVPDRDNDGIPDSLEVEGYVDVKKRFTLSFWIS 208
QY 61 NIEKKGLTKYKSSPEKWSASDPYSDFEYTGRIKNSPEARHPLVAAYPIVHVMEN 120
DB 209 NIEKKGLTKYKSSPEKWSASDPYSDFEYTGRIKNSPEARHPLVAAYPIVHVMEN 268
QY 121 IILSKNEDOSTQNTDSETRTISKNTSRTHTSEVHGNAEVHANTSRHTHTSEVHNAE 180
DB 269 IILSKNEDOSTQNTDSETRTISKNTSRTHTSEVHGNAEVHANTSRHTHTSEVHNAE 328
QY 181 VHAVALDHSLSLAGEPTMAETMGLNTADTARLANIIRYVMTGTAPIYVNLPTSLVGLKN 240
DB 329 SSTVALDHSLSLAGEPTMAETMGLNTADTARLANIIRYVMTGTAPIYVNLPTSLVGLKN 388
QY 241 QTLATIKAKENQSLIAPNNYPSKNLADIALNAODFSSPTITMNYNOFLELEKTKOL 300
DB 389 QTLATIKAKENQSLIAPNNYPSKNLADIALNAODFSSPTITMNYNOFLELEKTKOL 448
QY 301 RLTDQVYGNIAIYVNFENGVRVDTGSMNSEVLPOIETTARIIFENKDLNLVERRIAAY 360
DB 449 RLTDQVYGNIAIYVNFENGVRVDTGSMNSEVLPOIETTARIIFENKDLNLVERRIAAY 508
QY 361 NPSDPLETTKPDMLTKALKIAFGFNPNGNLOYOGKDITEFDENFPOQTSQNKKNOLAE 420
DB 509 NPSDPLETTKPDMLTKALKIAFGFNPNGNLOYOGKDITEFDENFPOQTSQNKKNOLAE 568
QY 421 LNAFTNIYTVLDKIKLAKNMNILLRDKRFHYDRNNIAVGADESYYKEAHREYINSTEGLL 480
DB 569 LNAFTNIYTVLDKIKLAKNMNILLRDKRFHYDRNNIAVGADESYYKEAHREYINSTEGLL 628
QY 481 LNTDKDIRKILSGYIYEIEDTEGLEKVINRYDMLNTSSLRQDGKTFIDPKKYNDKPLLY 540
DB 629 LNTDKDIRKILSGYIYEIEDTEGLEKVINRYDMLNTSSLRQDGKTFIDPKKYNDKPLLY 688
QY 541 ISNPNKVNYYAATKENT 558
DB 689 ISNPNKVNYYAATKENT 706

RESULT 7
US-08-082-849B-31
Sequence 31, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Kilmpel, Kurt R.
APPLICANT: Airoa, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
TITLE OF INVENTION: Related Methods
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-31

Query Match 89.9%; Score 2581; DB 1; Length 719;
Best Local Similarity 90.9%; Pred. No. 5.7e-199;
Matches 511; Conservative 6; Mismatches 21; Indels 24; Gaps 2;

QY 1 DNLQPELKOKSSNS---KKRSTASGPTVPDRDNDGIPDSLEVEGYVDVKKRFTLS 56
DB 149 DNLQPELKOKSSNSNTATIMQGNFLOGPTVPDRDNDGIPDSLEVEGYVDVKKRFTLS 208
QY 57 PMISNIEKKGLTKYKSSPEKWSASDPYSDFEYTGRIKNSPEARHPLVAAYPIVHV 116
DB 209 PMISNIEKKGLTKYKSSPEKWSASDPYSDFEYTGRIKNSPEARHPLVAAYPIVHV 268
QY 117 DMENIILSKNEDOSTQNTDSETRTISKNTSRTHTSEVHGNAEVHANTSRHTHTSEVH 176
DB 269 DMENIILSKNEDOSTQNTDSETRTISKNTSRTHTSEVHGNAEVHANTSRHTHTSEVH 328
QY 177 GNAEVHVAIDHSLSLAGEPTMAETMGLNTADTARLANIIRYVMTGTAPIYVNLPTSLV 236
DB 329 SNSNSSTVALDHSLSLAGEPTMAETMGLNTADTARLANIIRYVMTGTAPIYVNLPTSLV 388
QY 237 LGRNQTATIKAKENQSLIAPNNYPSKNLADIALNAODFSSPTITMNYNOFLELEK 296
DB 389 LGRNQTATIKAKENQSLIAPNNYPSKNLADIALNAODFSSPTITMNYNOFLELEK 439
QY 297 TKQLRLDTQVYGNIAIYVNFENGVRVDTGSMNSEVLPOIETTARIIFENKDLNLVERR 356
DB 440 -----YGNIAIYVNFENGVRVDTGSMNSEVLPOIETTARIIFENKDLNLVERR 488
QY 357 IAAVNPSPLETTKPDMLTKALKIAFGFNPNGNLOYOGKDITEFDENFPOQTSQNKKN 416
DB 489 IAAVNPSPLETTKPDMLTKALKIAFGFNPNGNLOYOGKDITEFDENFPOQTSQNKKN 548
QY 417 QLAELNATNIYTVLDKIKLAKNMNILLRDKRFHYDRNNIAVGADESYYKEAHREYINSTE 476
DB 549 QLAELNATNIYTVLDKIKLAKNMNILLRDKRFHYDRNNIAVGADESYYKEAHREYINSTE 608
QY 477 EGLLNTDKDIRKILSGYIYEIEDTEGLEKVINRYDMLNTSSLRQDGKTFIDPKKYNDK 536
DB 609 EGLLNTDKDIRKILSGYIYEIEDTEGLEKVINRYDMLNTSSLRQDGKTFIDPKKYNDK 668

QY 537 LPLYSNPNKYKVVYAVTKENT 558
| | | | |
Db 669 LPLYSNPNKYKVVYAVTKENT 690

RESULT 8

PCT-US94-01624-31
; Sequence 31, Application PC/TUS9401624
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Aroca, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
; STREET: Stewart Street Tower, 20th Floor, One Market
; STREET: Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01624
; FILING DATE: June 25, 1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-115
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-01624-31

Query Match 89.9%; Score 2581; DB 5; Length 719;
Best Local Similarity 90.9%; Pred. No. 5.7e-199;
Matches 511; Conservative 6; Mismatches 21; Indels 24; Gaps 2;

QY 1 DNLQPELKQSSNS---RKRSTSAQPTVPDRNDGIPDSLEVEGTVVYKKNRTPLS 56
| | | | |
Db 149 DNLQPELKQSSNSATIMQGNFLQGPYVDRNDSDIPSLSEVGEYTVDKNKRTEPLS 208
QY 57 PWISNIHEKGLTKYKSSPEKMTASDPYSDPEKTYGRIDKNVSEARHPLVAAPYIVH 116
| | | | |
Db 209 PWISNIHEKGLTKYKSSPEKMTASDPYSDPEKTYGRIDKNVSEARHPLVAAPYIVH 268
QY 117 DMENIILSKNEDOSTONTDSETRTTSKNTSTSRHTTSEVHGNAEVAHANTSTSRHTTSEVH 176
| | | | |
Db 269 DMENIILSKNEDOSTONTDSETRTTSKNTSTSRHTTSEVHGNAEVAHANTSTSRHTTSEVH 328
QY 177 GNAEYHAAVIAIDHSLSLAERTMAETMGUNTADTARLANIRVYNGTAPIYVLTPTSLV 236
| | | | |
Db 329 SNSNSTVAIDHSLSLAERTMAETMGUNTADTARLANIRVYNGTAPIYVLTPTSLV 388
QY 237 LGRNQTATITAKENQLSQIILAPNNYVPSKNLAPIALNAODEFSPTITMNNQGLELEK 296
| | | | |
Db 389 LGRNQTATITAKENQLSQIILAPNNYVPSKNLAPIALNAODEFSPTITMNNQGLELEK 439

QY 297 TKQLRLDTQVYGNATYVNFENGRRVVDTSNMSEVLPQIQETTARIIFNGKDLNIVER 356
| | | | |
Db 440 -----YGNATYVNFENGRRVVDTSNMSEVLPQIQETTARIIFNGKDLNIVER 488
QY 357 IAAVNPSPLETTKPDMTLKEALKIAFGNEPENGLOQOGKDIETEDPNDQOTSQNTKN 416
| | | | |
Db 489 IAAVNPSPLETTKPDMTLKEALKIAFGNEPENGLOQOGKDIETEDPNDQOTSQNTKN 548
QY 417 QLAELNATNYTVLQKIKINAKMNLIRDKRPHYDRNNIAGVADSVYEAHREYINNST 476
| | | | |
Db 549 QLAELNATNYTVLQKIKINAKMNLIRDKRPHYDRNNIAGVADSVYEAHREYINNST 608
QY 477 ECLLINIDKIRKILSGYIVIEIEDEGLKEVINDRYDMLNISLRDQKTFIDFKKYNDK 536
| | | | |
Db 609 ECLLINIDKIRKILSGYIVIEIEDEGLKEVINDRYDMLNISLRDQKTFIDFKKYNDK 668

RESULT 9

US-08-960-780-32
; Sequence 32, Application US/08960780
; Patent No. 6204435
; GENERAL INFORMATION:
; APPLICANT: Felleison, Jerald S.
; APPLICANT: Schepf, H. Ernest
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Stockhoff, Brian A.
; APPLICANT: Schmeltz, James
; APPLICANT: Loewer, David
; APPLICANT: Dullum, Charles Joseph
; APPLICANT: Muller-cohn, Judy
; APPLICANT: Stamp, Lisa
; TITLE OF INVENTION: No. 6204435el Pesticidal Toxins and Nucleotide
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/960,780
; FILING DATE: 30-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,848
; FILING DATE: 30-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA-708
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 881 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: 177C8
US-08-960-780-32

Query Match	Score	DB 4	Length	881
20.68	591			

```

0Y      1 DNLQLELQKQSS-----NSRKKRSTSAAPYPDRDNDGIPDSLEVEGYTD 47
Db      186 DELNRPENKKESEQEFLLAPSKINLFTQKKMKREIDED--TDRDGDSIPLMEENGTYI- 241
0Y      48 VKNKRTEFLSPINISIHKEKGLTKYKSSPEKMSKSTADDPDEPEKVTGIDKNVSPEARHPL 107
Db      242 ----QNRLLAVKWDSDL-ASKGTTKYVSNPPLSESHYGDPTDYDEKRAARDLSNAKETFPNPL 297
0Y      108 VAAYPPIVHDMENITILSKNEDOSTONTDSETRITSKNSTSRHTSEVHGNAEYHANTST 167
Db      298 VAAEPVNVSMKEVILSPENULS-----NSVSHSSTNW 331
0Y      168 SRRTSEVHGNAEYH-----AVADHSLSLAGERTMAETMG---LNTADTARLAN 215
Db      332 SYTTEGASVAGIGPKGISFGVSVNTOHSEVTAOE--NGTSTGNTSQFMTASAGYLNAN 389
0Y      216 IRYVNTGAPLYYVNLPTTSLVLGKNOTLATIKAKENQLOIILAPNNYPSKNLAPIALNA 275
Db      390 VRIYVNTGALITYDYKPTTSPYL-NNDTIATITAKSNSLTALNISGESYPAKGONGIATIS 448
0Y      276 QODESSTPTITANNYQFLELEKTKOLRLDTPQVYGNATYFENGGRVAVDGSNMSEVLPQ 335
Db      449 MDDENSHPTLTKKQYVDLNLNKKRPMLETNOTDG---VYKIKDPHGNIVYGGEMNGVIQ 505
0Y      336 IOETTARIIFNGKRLNLYERIAVANSDDPLETKKPRMTLKEAKIAF--GFNEPKNQLO 353
Db      506 IKAATASTIYVDGE-RAVEKKVAAKDIENBEDKT-PSLTLTKDAKLKLYPDEIKETIEELLY 563
0Y      394 YQGRDITEFDF--NPDQOTSQINQOLAEL-----NATNTIYTVLDKIKLAKMNIILRDK 446
Db      564 YKKNPIYESSWYTLDEMTAEVIGKOLNDPTGKRGKVDXSHLYDV----KLTPKMWITJK-L 618
0Y      447 RFHYDRNNIAVGADESVYKKAHREYVINSSTEG-----LLIIND-----KDIR 488
Db      619 SILYDN--AESNDSNIGCKWNTNTIVSGVGNKKQKQYSSNPDANLLTNTDAOEKLNMRD 675
0Y      489 KILSGYIIEIDTEGLKEVINDRYDMLNIS-SLRQDOKTFIDFKKYNDKLPIYSNPNYK 547
Db      676 YYISLYMKSEKNTQCEITIDELIYPTTKYVNVKNQYKRLDIITAHNK-----SNPISS 730
0Y      548 VNV 550
Db      731 IHI 733

RESULT 10
US-09-073-898-32
US-Sequence 32, Application US/09073898
US-Patent No. 6242669
US-GENERAL INFORMATION:
US-APPLICANT: Feltelson, Jerald S.
US-APPLICANT: Schaeff, H. Ernest
US-APPLICANT: Naray, Kenneth E.
US-APPLICANT: Stockhoff, Brian A.
US-APPLICANT: Schmeltz, James
US-APPLICANT: Loewer, David
US-APPLICANT: Dullum, Charles Joseph
US-APPLICANT: Muller-Cohn, Judy
US-APPLICANT: Stamp, Lisa
US-APPLICANT: Morrill, George
US-APPLICANT: Finstad-lee, Stacey
US-TITLE OF INVENTION: Sequences 6242669el Pesticidal Toxins and Nucleotide
US-Sequences Which Encode These Toxins
US-NUMBER OF SEQUENCES: 144
US-CORRESPONDENCE ADDRESS:

```

ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville

CLASSIFICATION:

APPLICATION NUMBER: US 60/029,848

APPLICATION NUMBER: US 08/960,780

NAME: Sanders, Jay M.

TELEPHONE: 352-375-8100 ; TELECOMMUNICATION INFORMATION: ;

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 881 amino acids
;

```

INDIVIDUAL ISOLATE: PS177C8
US-09-073-898-32

Query Match	20.6%	Score 591	DB 4	Length 881
Best Local Similarity	29.9%	Pred. No. 4.3e-39		
Matches 180; Conservative	97	Mismatches 218	Indels 108	Gaps 22

[illegible]

```
Db 564 YAKKPIYESVWTFYIDENAKETVKQKLDVTGKFKVDVSHLYDV-----KLTPrMANTIK-L 618
Qy 447 RHHYRNNIAGADSVYKAEHREVINSSTEG-----LLNID-----KDIR 488
Db 619 SLTYDN---AESNDNSIGKWTNTIVSGGNNGKKOYSSNPNPDLNTDADEKLNKND 675
Qy 489 KILSGYIEIEETEGLEKYINDRYDMLNIS-SLRDQGTFTIDFKKYNDKLPYISNPYK 547
Db 676 YIISLYMSEKNTQCEITIDGELIPIITTKTVNVNKNDRKRLDIIAHNIK-----SNPISS 730
Qy 548 VNV 550
Db 731 IHI 733

RESULT 11
US-08-471-033-5
; Sequence 5, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostlichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESS: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.308
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471.033
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: P-40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-471-033-5
```

```
Query Match 20.5%; Score 589; DB 1; Length 884;
Best Local Similarity 30.2%; Pred. No. 6,3e-39;
Matches 180; Conservative 94; Mismatches 215; Indels 108; Gaps 22;

Qy 1 DNLQIPDLKOKSS-----NSRKKRSTAGPIVPRDNDGIPDLSLEVGYVD 47
Db 189 DELRNEEFNKKRQSEFLAKPSKINLFTQMKREIDED---TPTDDSDIPDLWEENGTYI- 244
Qy 48 YNKRRTFLSPWISNIHEKGLTKFYKSPPEKMTASDPYSDFEYVGRIDKNVSPARPL 107
Db 245 ---QNRIVAKWDDSL-ASGQYTKFYVSNPLESHTVGDPTIDYKARDDLSNAKETFNP 300
Qy 108 VAAVPIVHVDENIILSKNEDOSTONTDSEFTTKNTSTSTHTMSEVHNAEVAHNTST 167
Db 301 VAAFPVSVNYSMEKVILSPENLS-----NSVESHSSTNW 334
Qy 168 SRTHTSEVHNAEVAH-----AAIDHSLSLAGERTWAEWNG-----LNTADTARLANN 215
Db 335 STYNTGASVNEGIGPKGISFGVSVNYQHSETVAGE--WGTSJGNTSOFNTASAGYLANN 392
Qy 216 IRYVMTGTAPIYVNLPTSLVGLKQNTLATIKAKENQSLQILAPNNYPSKNLAFIALNA 275
Db 393 VAYNVNGGAIYDVAKPTTSFVL--NNDTIATITAKSNSTALNISPESEYPKKONGIATIS 451
Qy 276 ODEFSSTPTNNYNOFLEKTKQLRLDQYGNIAFYNEFNGRVNRVDGSMSEVLPO 335
Db 452 MDFFNSHPTTLNKKQVDNLLNKKPMULETNTQDQ---YKIKDTGNIYTGGEWNGVIOQ 508
Qy 336 IOETARIITFNCKDLINVERIAAVNPSDPLETTKPDMTIKALKATIAF--GENENGNIQ 393
Db 509 IAKTRASITVDGE-RVAKERYAAKQVNEPEDKT-PSLTALKALKLSVPDEIKETEGELY 566
Qy 394 YQKDIITEFD--NFDQOTSONIKQLAEL-----NATNIYVLDKIKLNAKMNILIRDK 446
Db 567 YNKRPIYESVWTFYIDENAKETVKQKLDVTGKFKVDVSHLYDV-----KLTPrMANTIK-L 621
Qy 447 RHHYRNNIAGADSVYKAEHREVINSSTEG-----LLNID-----KDIR 488
Db 622 SLTYDN---AESNDNSIGKWTNTIVSGGNNGKKOYSSNPNPDLNTDADEKLNKND 678
Qy 489 KILSGYIEIEETEGLEKYINDRYDMLNIS-SLRDQGTFTIDFKKYNDKLPYISNP 544
Db 679 YIISLYMSEKNTQCEITIDGELIPIITTKTVNVNKNDRKRLDIIAHNIK-----SNP 730

RESULT 12
US-08-471-044-5
; Sequence 5, Application US/08471044
; Patent No. 5840868
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostlichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESS: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.308
```


CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,044
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-471-044-5

Query Match 20.5%; Score 589; DB 2; Length 884;
Best Local Similarity 30.2%; Pred. No. 6,3e-39;

Matches 180; Conservative 94; Mismatches 215; Indels 108; Gaps 22;

QY 1 DNQLPELKOKSS-----NSRKRSTSGAPTPVDRDNDGIPDSLEVEGYTVD 47
DB 189 DELRNPEFNKKEOEFLLAKSKINLTQKKMKREIDED---TDDGDSIPDLMEENGTTI- 244
QY 48 VKNKRTFLSPWISNIHEKGLTKYKSSPEKWSSTASDPYSDPEKVTGRIDKNVSPPEARHPL 107
DB 245 ---QNRIVAKWDDSL-ASKGYTKFVSNPLESHYVGDPYTYDEKAAARLDLSNAKETNPPL 300
QY 108 VAAPIYVHDMENIILSKNDOSTONTDSETRITISKNTSTSRHTSEVHGNAEYHANTST 167
DB 301 VAAFPVSNVSMEXVILIPNENLS-----NSVSHSSTNW 334
QY 168 SRHTSEVHGNAEYH-----AVADHSLSLAGERTAAETWG---LNTADTARLNNAN 215
DB 335 STYNTGASVEAGIGPKGISFGVSNVQHSSETVAQE--WGTSTGNTSQFNTASAGYLNAN 392
QY 216 IRYVNTGTAPIVNVLPSTSLVLGKNOTLATIKAKENOLSOILAPNNYPSKNLAPIALNA 275
DB 393 VRYNNTGTGALYDVKPTTSFVL--NNDTIATITAKSNSTALNISGEEYPRKGGNGIATIS 451
QY 276 QDDFSSTPTMANQLELEKTKQLRLDTQVYGNLATYFENGRAVAVDTGSMSEVLPQ 335
DB 452 MDEFNSHPITLANKQVDNLNNKPMLETTNOTG---VYKIKDTHGNIIVGGEENGVIYQ 508
QY 336 IOETRTITFNGKDLNVERIAAVNPSPDLETTTKPDMTEKALKIAF--GFNEPNNLQ 393
DB 509 IKAKTISIIVDGE--RAEKRVAAKTDENPEDKT--PSLTIKDLKLSYDPPEIKETIBELLY 566
QY 394 YOGKDIETEDF--NPDQOSTONIKNOIAEL-----NATNIYTVLDKIKLAKMNIILDRK 446
DB 567 YKRPPIYESSVMTYDENTAKETKQDLNDTGTGFKVDYSHLYD---KLTPKNAVITIK-L 621
QY 447 RFIYDRNNINAVGDESVEYKAHREVINSSTEG-----LLINID-----KDIR 488
DB 622 SILYDN---AESNDNSIGKWTNTNIVSGCNGNKKOYSSNNPDANLLTINTDAOEKLNKRD 678
QY 489 KILSGIVEIEDEGLKEVINDRYDMLNIS--SLRQCKTFIDKKYNDKLPVYISNP 544

Db 679 YVLSLVKSKENKNOCELTITGEIPIPTTKYVNAKKDKVNRDLIAHNIK-----SNP 730

RESULT 13
US-08-463-483A-5
Sequence 5, Application US/08463483A
Patent No. 5849870
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Koziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalin M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
TITLE OF INVENTION: No. 5849870el Pesticidal proteins and strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
City: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.308
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murtry
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-463-483A-5

Query Match 20.5%; Score 589; DB 2; Length 884;
Best Local Similarity 30.2%; Pred. No. 6,3e-39;

Matches 180; Conservative 94; Mismatches 215; Indels 108; Gaps 22;

QY 1 DNQLPELKOKSS-----NSRKRSTSGAPTPVDRDNDGIPDSLEVEGYTVD 47
DB 189 DELRNPEFNKKEOEFLLAKSKINLTQKKMKREIDED---TDDGDSIPDLMEENGTTI- 244
QY 48 VKNKRTFLSPWISNIHEKGLTKYKSSPEKWSSTASDPYSDPEKVTGRIDKNVSPPEARHPL 107
DB 245 ---QNRIVAKWDDSL-ASKGYTKFVSNPLESHYVGDPYTYDEKAAARLDLSNAKETNPPL 300

APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalin M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5872212el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5872212artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,566B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meligs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ. ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-470-566B-5

Query Match 20.5%; Score 589; DB 2; Length 884;
Best local Similarity 30.2%; Pred. No. 6,3e-39;
Matches 180; Conservative 94; Mismatches 215; Indels 108; Gaps 22;

QY 1 DNLQLELKOKSS-----NSRRKSTSGPTVPRDRNDGIPDSLEVEGYTD 47
DB 189 DELRNPENFKKESQEFLLAKPSKINLFTQKKREIDED---TDTGDSIPDLMEENGTYI- 244
QY 48 VKKKRTFLSWINIEHKKLTLYKSSPEKMWSTASDPYSDPEKYTGRIIDKNVSPFARHPL 107
DB 245 ---QNRIAVMWDSL-ASKGYTKFVSNPLSHVGVGPDYTDYERKAARDLDSNAKEFTNPPL 300
QY 108 VAAYPYIVHDMENILSKNEDOSTONTDSETRTISKNTSTSRHTSEVHGNAEYHANTST 167
DB 301 VAAFPVNVSMKVIILSPNMLS-----NSVESHSSTNW 334
QY 168 SRHTSEVHGNAEVH-----AVAIHSLSLAGERTMAETMG---LNTADTARLNNAN 215
DB 335 SYNTTEGASVEAGIGPKGISFVGSVNYQHSETVAQE--MCTSTGNTSQPNTASAGYLNAN 392

QY 216 IRVYNTGTAPIYVNLPTTSIVLGKNOTLATIKAKENOLSOILAPNNYPSKNLAPIALNA 275
DB 393 VRYNNVGTGAIYDVKPTTSYVL--NNDIATITAKSNSTALNISGESYPRKGGONGIATIS 451
QY 276 QDDFSSPTITMANNQFLEKTKQLRLDTQVYGNIAITYFENGVRVVDYTGSMNSEVLPO 335
DB 452 MDPFNSHPITLNNKQVNDLNNKPMLETNOTDG---VYKIKDTHGNIYVGGEWNGV10Q 508
QY 336 IOETTAIINGKDLNIVERIRIAAVNPSDPLETTKPMTLKALKIAF--GFNPENGNLQ 393
DB 509 IKAKTASIIIVDGE-RVAEKRVAAKDYENPEDKT-PSLTLKDALKLSYPOEIKETEGLLY 566
QY 394 YOGKDIREFDF--NFDQOTSQNIKNOLAEL-----NATNIYTVYDKTKLAKNNILIRDK 446
DB 567 YKKNPIYESSVMYTLDENTAKYVKQLNDITTGAKKDYSHLYDV---KLTPKMNVTIK-L 621
QY 447 RFHYDRNNIAVGADSEVYKKAHREVINSSTEG-----LLNID-----KDIR 488
DB 622 SILYDN---AESNDNSIGKMTNTNIVSGNNGKKQYSSNNDAN/LTNTDAQEKLNKRD 678
QY 489 KIISGYIVLEIEDTEGLKEVINDRYDMLNIS-SLRQDKTFIDFKKYDKLPLYISNP 544
DB 679 YVISTLYMKSEKNTQCEITIDGELYPIITTKYVNVNKKDYKRLDIHANIK-----SNP 730

Search completed: December 2, 2001, 13:50:01
Job time: 241 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:54:38 ; Search time 132.69 seconds
(without alignments)
842.204 Million cell updates/sec

Title: US-09-747-521-4
Perfect score: 3913
Sequence: 1 MKRKKVLPMLALSTLVSS.....TSTNGIKILIFSKGYEIG 764

Scoring table: BROSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPTREMBL_17:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3797	97.0	764	2	09RQ02
2	3793	96.9	764	2	09KH69
3	3785	96.7	764	2	09F5R7
4	900	23.0	876	2	032739
5	898	22.9	876	2	09KH41
6	886	22.6	879	2	006498
7	883.5	22.6	875	2	046221
8	798	20.4	721	2	086171
9	235.5	6.0	204	2	09X377
10	205.5	5.3	4688	2	09PQ08
11	199.5	5.1	1387	5	09G276
12	197	5.0	2647	5	09G4X0
13	195.5	5.0	3254	5	09BK45
14	193.5	4.9	3130	5	09BK46
15	193.5	4.9	4919	2	09ZHL0
16	193	4.9	1072	2	09CF64
17	190	4.9	1302	2	049547
18	189	4.8	6713	2	099054
19	188.5	4.8	2771	5	026216

20	186	4.8	2178	2	046149	046149	clostridium
21	182.5	4.7	1193	2	045914	045914	clostridium
22	182	4.7	2747	5	09BX9	09BX9	plasmidium
23	180.5	4.6	149	2	09RM77	09RM77	clostridium
24	180	4.6	1051	2	049524	049524	mycoplasma
25	180	4.6	2269	5	026223	026223	plasmidium
26	179.5	4.6	962	2	049546	049546	mycoplasma
27	179.5	4.6	1365	2	049525	049525	mycoplasma
28	179	4.6	769	2	099Y70	099Y70	staphylococ
29	178.5	4.6	149	2	09RM78	09RM78	clostridium
30	178.5	4.6	1193	2	P71107	P71107	clostridium
31	177	4.5	1315	2	086488	086488	staphylococ
32	177	4.5	1837	3	074424	074424	schizosacch
33	177	4.5	3724	5	077320	077320	plasmidium
34	176.5	4.5	149	2	09RM80	09RM80	clostridium
35	176	4.5	1621	5	024984	024984	giardia lam
36	175.5	4.5	1516	5	096154	096154	plasmidium
37	175	4.5	1176	2	09F231	09F231	helicobacte
38	175	4.5	1939	5	025662	025662	plasmidium
39	174.5	4.5	149	2	09RM79	09RM79	helicobacte
40	174	4.4	2529	2	025579	025579	plasmidium
41	173.5	4.4	1344	2	049545	049545	mycoplasma
42	173	4.4	1230	2	025772	025772	helicobacte
43	171.5	4.4	821	3	008581	008581	saccharomyc
44	171.5	4.4	1098	5	09C275	09C275	plasmidium
45	170	4.3	1152	4	092603	092603	homo sapien

ALIGNMENTS

RESULT 1
ID 09RQ02; PRELIMINARY; PRT; 764 AA.
AC 09RQ02;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PX01-110 (PROTECTIVE ANTIGEN).
GN PAg.
OS Bacillus anthracis.
OC Plasmid virulence plasmid PX01, and Plasmid PX01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPERNE; PLASMID=VIRULENCE PLASMID PX01;
RX MEDLINE=99445483; PubMed=10515943;
RA Okinaka R.T., Cloud K., Hampton O., Hoffmaster A.R., Hill K.K.,
RA Keim P., Koehler T.M., Lamke G., Kumano S., Mahillon J., Manter D.,
RA Martinez Y., Rieke D., Svensson R., Jackson P.J.,
RT "Sequence and organization of px01, the large Bacillus anthracis
RT plasmid harboring the Anthrax toxin genes.";
RL J. Bacteriol. 181:6509-6515(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=33, 28, AND BA1035; PLASMID=PX01;
RX MEDLINE=99214082; PubMed=10197996;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.,
RT "Genetic diversity in the protective antigen gene of Bacillus
RT anthracis.";
RL J. Bacteriol. 181:2358-2362(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=33, 28, AND BA1035; PLASMID=PX01;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.,
RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF065404; AAD32414.1;
DR EMBL: AF030678; AAG24449.1;
DR EMBL: AF306778; AAG24449.1;
DR EMBL: AF306779; AAG24447.1;
DR EMBL: AF306780; AAG24448.1;

DR HSP: P13423; JACC.
DR InterPro: IPR003896; Binary_toxB.
DR PRINTS: PR01391; BINARYTOXINB.
KW Plasmid.
SO SEQUENCE 764 AA; 85810 MW; 3AE1EBF48FAA03F CMC64;

Query Match 97.0%; Score 3797; DB 2; Length 764;
Best Local Similarity 97.5%; Pred. No. 3.3e-208;
Matches 745; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

```
QY 1 MKKRVLLPMAALSTIVSSGNGNEVIAEYKQENRLNESSSSGGLGYFSDLNQA 60
DB 1 MKKRVLLPMAALSTIVSSGNGNEVIAEYKQENRLNESSSSGGLGYFSDLNQA 60
QY 61 PMVYTSSTGDLSTPSSSELENIPSENOYFQSAIMSGFTKVKKSDERTYATSDNHTVMV 120
DB 61 PMVYTSSTGDLSTPSSSELENIPSENOYFQSAIMSGFTKVKKSDERTYATSDNHTVMV 120
QY 121 DDQEVINKASNSNKTRELEKGRLYQIKIYORENPTKEGLDFKLWYTDSONKKEVISSDNL 180
DB 121 DDQEVINKASNSNKTRELEKGRLYQIKIYORENPTKEGLDFKLWYTDSONKKEVISSDNL 180
QY 181 QLPETKQSSNRKRRSTASAGTPVDRNDGIPDSELEEGYVDVKNKRTFLSPWISNH 240
DB 181 QLPETKQSSNRKRRSTASAGTPVDRNDGIPDSELEEGYVDVKNKRTFLSPWISNH 240
QY 241 EKKGLTKKSSPEKMTASDPYSDPEKTYGRIDKNVSEARHPLVAAPYIVHDMENITL 300
DB 241 EKKGLTKKSSPEKMTASDPYSDPEKTYGRIDKNVSEARHPLVAAPYIVHDMENITL 300
QY 301 SKNEDQSTQNTDSETRTISKNTSTSRTHTSEVHGNAEVAHANTSTSRTHTSEVHGNAEVAHA 360
DB 301 SKNEDQSTQNTDSETRTISKNTSTSRTHTSEVHGNAEVAHANTSTSRTHTSEVHGNAEVAHA 360
QY 361 VAIDHSLSLAGERTAEETMGNTADTARLANIRVNTGTAPYVAVLPTTSLVLCGNQTL 420
DB 361 VAIDHSLSLAGERTAEETMGNTADTARLANIRVNTGTAPYVAVLPTTSLVLCGNQTL 420
QY 421 ATIRAKENQSOIILAPNNYPSKNLAPIALNAODFSSPTITMANNQLELEKTYQQLRD 480
DB 421 ATIRAKENQSOIILAPNNYPSKNLAPIALNAODFSSPTITMANNQLELEKTYQQLRD 480
QY 481 TDQVYGNATYTNFENGRAVRVDTGSNMSEVLPOIETTAIIIFNGKDLNVERRIAANVPS 540
DB 481 TDQVYGNATYTNFENGRAVRVDTGSNMSEVLPOIETTAIIIFNGKDLNVERRIAANVPS 540
QY 541 DPLETTKPDMLKEALKIAFGNPNNGMLQYQKDIETFDNFDDQTSQNTKNOIAELNA 600
DB 541 DPLETTKPDMLKEALKIAFGNPNNGMLQYQKDIETFDNFDDQTSQNTKNOIAELNA 600
QY 601 TNYIVLDDIKILNAKMNILIRDKRHYDRNNIAGADESVYKEAAREVINSSTEGILLNI 660
DB 601 TNYIVLDDIKILNAKMNILIRDKRHYDRNNIAGADESVYKEAAREVINSSTEGILLNI 660
QY 661 DKDIRKILSGYIVIEIDTEGLKEVINDRYDMLNISLRQDKGTIDFKKYNDKLPYLYSN 720
DB 661 DKDIRKILSGYIVIEIDTEGLKEVINDRYDMLNISLRQDKGTIDFKKYNDKLPYLYSN 720
QY 721 PNYKVNVAATKENTINPSENGDSTNGIKKILFSSKKGYEIG 764
DB 721 PNYKVNVAATKENTINPSENGDSTNGIKKILFSSKKGYEIG 764
```

RESULT 2

Q9KH69 PRELIMINARY; PRT; 764 AA.
AC Q9KH69;
DT 01-OCT-2000 (TReMBLrel. 15; Created)
DT 01-OCT-2000 (TReMBLrel. 15; Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17; Last annotation update)
DE PROTECTIVE ANTIGEN.
GN PAGE OR PAG.

OS Bacillus anthracis.
OG Plasmid pX01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;

```
RA [1]
RA SEQUENCE FROM N.A.
RA STRAIN-V770-NP1-R, ATCC14185; PLASMID=PX01;
RA MEDLINE=20359347; PubMed=10899854;
RA Cohen S., Mendelson I., Altboum Z., Kobler D., Elhanany E., Bino T.,
RA Leitner M., Inbar I., Rosenberg H., Gozes Y., Barak R., Fisher M.,
RA Kronman C., Velan B., Shafterman A.;
RA "Attenuated nontoxinogenic and nonencapsulated recombinant Bacillus
RA anthracis spore vaccines protect against anthrax.";
RA Infect. Immun. 68:4549-4558(2000).
```

```
RL [2]
RL SEQUENCE FROM N.A.
RL PLASMID=PX01.
RL MEDLINE=99214082; PubMed=10197996;
RL Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RL "Genetic diversity in the protective antigen gene of Bacillus
RL anthracis.";
RL J. Bacteriol. 181:2358-2362(1999).
```

```
RL [3]
RL SEQUENCE FROM N.A.
RL PLASMID=PX01.
RL Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF268967; AAF6457.1; -
RL EMBL: AF306782; AAG24450.1; -
RL InterPro: IPR003896; Binary_toxB.
```

DR PRINTS: PR01391; BINARYTOXINB.
DR Plasmid.
SO SEQUENCE 764 AA; 85838 MW; B0DAFC1DCF83DAF4 CMC64;

Query Match 96.9%; Score 3793; DB 2; Length 764;
Best Local Similarity 97.4%; Pred. No. 5.6e-208;
Matches 744; Conservative 3; Mismatches 17; Indels 0; Gaps 0;

```
QY 1 MKKRVLLPMAALSTIVSSGNGNEVIAEYKQENRLNESSSSGGLGYFSDLNQA 60
DB 1 MKKRVLLPMAALSTIVSSGNGNEVIAEYKQENRLNESSSSGGLGYFSDLNQA 60
QY 61 PMVYTSSTGDLSTPSSSELENIPSENOYFQSAIMSGFTKVKKSDERTYATSDNHTVMV 120
DB 61 PMVYTSSTGDLSTPSSSELENIPSENOYFQSAIMSGFTKVKKSDERTYATSDNHTVMV 120
QY 121 DDQEVINKASNSNKTRELEKGRLYQIKIYORENPTKEGLDFKLWYTDSONKKEVISSDNL 180
DB 121 DDQEVINKASNSNKTRELEKGRLYQIKIYORENPTKEGLDFKLWYTDSONKKEVISSDNL 180
QY 181 QLPETKQSSNRKRRSTASAGTPVDRNDGIPDSELEEGYVDVKNKRTFLSPWISNH 240
DB 181 QLPETKQSSNRKRRSTASAGTPVDRNDGIPDSELEEGYVDVKNKRTFLSPWISNH 240
QY 241 EKKGLTKKSSPEKMTASDPYSDPEKTYGRIDKNVSEARHPLVAAPYIVHDMENITL 300
DB 241 EKKGLTKKSSPEKMTASDPYSDPEKTYGRIDKNVSEARHPLVAAPYIVHDMENITL 300
QY 301 SKNEDQSTQNTDSETRTISKNTSTSRTHTSEVHGNAEVAHANTSTSRTHTSEVHGNAEVAHA 360
DB 301 SKNEDQSTQNTDSETRTISKNTSTSRTHTSEVHGNAEVAHANTSTSRTHTSEVHGNAEVAHA 360
QY 361 VAIDHSLSLAGERTAEETMGNTADTARLANIRVNTGTAPYVAVLPTTSLVLCGNQTL 420
DB 361 VAIDHSLSLAGERTAEETMGNTADTARLANIRVNTGTAPYVAVLPTTSLVLCGNQTL 420
QY 421 ATIRAKENQSOIILAPNNYPSKNLAPIALNAODFSSPTITMANNQLELEKTYQQLRD 480
DB 421 ATIRAKENQSOIILAPNNYPSKNLAPIALNAODFSSPTITMANNQLELEKTYQQLRD 480
QY 481 TDQVYGNATYTNFENGRAVRVDTGSNMSEVLPOIETTAIIIFNGKDLNVERRIAANVPS 540
```


Db	5	MRNKKVSFLTLTAIVSOLALVPPYAQSTSN-----HSKKKEIYNEDILPNNGLMG	58
Qy	52	YFSDLNFOAPRWVTSSTTGDSTFSSSELEN-IPBENQFOSALHSGFIKAKKSDETFEAT	110
Db	59	YFSDHFEKDLKMAPIKGKLNLFPEKKYKGLDKDKSDVSRITKGIIFSKOGEGYTLST	118
Qy	111	SADNHNVTMMVADOEYINKASNSKRIEKGRLVQIKIOYOREN--PREKGLDFKLWTFDS	168
Db	119	DRDD-VLMQVNTTESTI---SMTLKVNKKKGEYIVRIELODKNLGSDIENSSPLWY-EL	173
Qy	169	QNKKEVISDDNLQDPELKOKSSNRKKRSTAGPTV-----PRDND	210
Db	174	DGMKKIIPPEENLPLRDY-----SNIEKD-----DPFIRNNPFDPKLMSDMEDDLDTDND	224
Qy	211	GIPDSLEVEGYTVDKKNKRTFLSPWISNIHEKKGLTYKSSPEKWSASDPYSDFEKVTG	270
Db	225	NIPSYERNNGTYI-----KDLIAVWEDSSFAE-QGYKKYVSNTLESNAGCPDYDEKASG	279
Qy	271	RIDKRVSEAAHPRIYAAPRIYHVMENIILSKNDSOTQNTDSTRTISAKNTSRRHTS	330
Db	280	SFDAIKTEADPPLAARPVIGVMEKLIISTNSHAST---DQKIVSRATNTSKESN	335
Qy	331	EVHGNAEV-----HANTSRTRTSEVHGNAEVAAVADHSLSLAGEPWAETMGNTA	384
Db	336	TAGSVAVNGCYNGGTANTTNYSHTTD-----NSRAVDS---NE-SMNGSLSTNGK	384
Qy	385	DTARLANIRYVNTGTAPRIYVNLPTTSLVLGKNOQTATIKAKENQSOILAPNNYPSKN	444
Db	385	ESAYINAVRYNNGTAPRMVKYPTPTNLVL-DGDTLSTIKAEQNIQNNLSPGDTYPKG	443
Qy	445	LAPITALNODPFSSPTPTMNNQPLEKTKQRLDLDYVGNATATYVFNENGRVRDTS	504
Db	444	LSPLALNTMDPFSSRLPIINTDQKLKLDAGQIKLETTOVSGNGTATN-SSGQI-VTEGN	501
Qy	505	NMSEVLPQIOETTARIIFNGKDLIVERRIAVNPSPLETTKPKDMLKEALAKIAEFNE	564
Db	502	SMSDVISQIDISAIISIIDTEN-ESYERRVYAKMLQDPEDKT-PELLTGAIEAKGAPATK	559
Qy	565	PNGNLQYQKQDITE-FDENFDQOTSONIKQNALNATNIVYLDKIKLNAMNILLRD	622
Db	560	KDGLLEFYNDIDESCVELLIPDDMTANKIKDSTLSDKKIYV---KLERGMNILLKT	615
Qy	623	KRFPH---DRNNIIVGADESIVKKAHREVINSSTEGL-----LLNDKDKRLKLSG	670
Db	616	PTYTFNEDDYNPNP--STWSVNNTTNDQGLQGSANKINGETKIKIPMSELSPKYKRYFSG	673
Qy	671	Y-----IVEI-----EDTEGL-----KEVINDRYDMLNIS-SLRDQKTFI	705
Db	674	YSKODPLTNSIIVTKAKKEKTDVLVPEQGTAKTSYEFETTERDSSMIETLLIGSGTYL	733
Qy	706	DFKKYND-KLPLYSINPNYKV-----NYVAVTKEINTIINPSENGDSTNGI	750
Db	734	DNLSTELTNSPPEILDEPEVKIPQDEIMDAHKIFYADLNFNPS-TGNITYINGM	786
RESULT	5		
Q9KH41		PRELIMINARY;	PRT; 876 AA.
AC	Q9KH41		
DT	01-OCT-2000 (Tremblrel. 15, Created)		
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)		
DT	01-JUN-2001 (Tremblrel. 17, Last annotation update)		
DE	CDTB.		
GN	CDBB.		
OS	Clostridium difficile.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;		
OC	Clostridium.		
OX	NCBI_TaxID=1496;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	STRAIN=CCUG 20309;		
RA	Chang S.Y., Song K.P.;		
RT	ADP-ribosylating Binary Toxin Genes of Clostridium difficile strain		

```

RT      CUG 20309."
RL      Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
FOR      EMBL; AF217119; AAF01761.1;
DR      InterPro; IPR003896; Binary-toxB.
DR      PRINTS; PRO1391; BINARYTOXB.
SO      SEQUENCE      876 AA;  98792 MM;  366D62F352E745A5 CRC64;

Query Match      22.9%; Score 998; DB 2; Length 876;
Best Local Similarity 32.0%; Pred. No. 4e-43;
Matches 267; Conservative 146; Mismatches 285; Indels 136; Gaps 36;

QY      1 MKKRVL--IPLMAL-STILV-----SSTGLVEYIOAEVQENMLLNSESSSOGCLCY 51
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      5 MNRKVLSEFLTALVALVSOALVYVYAQSTSN-----HSNKKKEIVNEDLLPNNGLMGY 58
QY      52 YFSDLNFOAPMVVTSSTGDLISIPSELEEN-IPSENOYFQSAIWGSGFLTKVKKSDRYFAT 110
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      59 YTFDHFQDLKLMALIKGNLKFEEKKVDKLLDKKSDVKSIRMTGRILIPSKOGEYTLST 118
QY      111 SADNHTVMVNDQOEYINKASNSNKKIRLEGRLYQIKIQOREN--PTEKGLDFKLYWDS 168
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      119 DRDD-VLQOVNTESTI-----SNTLKVMKKKREYKRIELODKNLGSDISPLNYW-EL 173
QY      169 QNRKVISSDNLQIPELQOKSSNSKKKSTAGRPV-----DNRND 210
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      174 DGMKKIIEENELFLDY---SNIEKD---DPEIPNNFFDPKIMSDWEDEDLDTDND 224
QY      211 GIPDSEVEGYTVADVKNKRTFLSPISNHEKKGLTKKSPKSKTASDPYDFEKYTG 270
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      225 NIPDSYERNGTYI---KDLIAVKHEDSPAE-QGKTKKYSNTLENTYADPTTIDEKASG 279
QY      271 RIDKNVSEARHPVLAAPYIVHVDENITILSKNEDOSTQNTDSETRTISKNTSTSRHTS 330
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      280 SFDKAIKTEARDPLVAAPYIVGMEKLIISTNEHAST---DQKTVSRATTNKTESN 335
QY      331 EVHGAEEV-----HANTSTSRTHSEVHGNAEVAIVADHSLSLAGERTAEYMGNTA 384
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      336 TAGVSVNNGYONGFTANVTNTNSHTTD-----NSTAVOOS--NGE-SWMTGLSINK 384
QY      385 DTARLNNANRYNVTGTAPLVNVLPTTSLVIGKNTQIATIKAKENOLSOILAPNNYPSKN 444
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      385 ESATINANVRYNTGTAPMYKYVPTTNLYL-DGDTLSITIKAOENIGNNLSPGDTPPKG 443
QY      445 LAPIALNADDESSPTITMANYNQLELEKTKQLRLDTDOYVGNITATYFENGREVRVDTGS 504
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      444 LSPALNTMDQESSRLIPNDQLKKLADGQIKLETTGYVSGNFCTKN--SSQI-VTEGN 501
QY      505 NMSEVLPQIQTETATIRNGKDLNVERIAAVNSDPLETTKPMDTLKEALKAFGENE 564
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      502 SMSDYSIOIDSISASIIIDTEN-ESYERVYPAKNLQDEDEKT-PELTIGEALEKAFGATK 559
QY      565 PNGNLQYOGKIDTE--PDPNFQOOTSQNLKNDLAEMLNTNTYTVDKIKLKNKMLILRD 622
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      560 KDGLLYENDIPIDESCEVLEIPDNTANKIKDSKLTLSDKKIYNV---KLEGMNILLKT 615
QY      623 KRFHY---DRNNIAGADESVVKEAHREYNSSTEGE-----LNLNDIKRILISG 670
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      616 PLYFNPEDYNNYR--SWMSNVNTNKGLOGSANKLNGETIKIKIPMSLAKRYKRYVSG 673
QY      674 YSKDPLTNSIIVKIKAKEEKTQYLVPEQGYTKFSEFETTEKDSSEIETLIGSGTYL.733
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      706 DFKKYND--KLPLYISNPYK-----VNVYAVVKEKNTIIPSENGDPTSGI 750
DB      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      734 DNLSTELNSTPEIIDPEVKIPTQEIIDAHKIIFADLNPNPS--TGNTYINGM 786

```


DT 01-JUL-1997 (TReMBLrel. 04, Created)
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE SB COMPONENT.
 GN SRS.
 OS Clostridium spiroforme.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes.
 OX NCBI_TaxID=29348;
 RN [1]
 RP SQUENCE FROM N.A.
 RC STRAIN-CS246;
 RA Gilbert M., Perelle S., Daube G., Popoff M.R.;
 RL Syst. Appl. Microbiol. 20:337-347(1997).
 DR EMBL; X97969; CAA66612.1; -
 DR HSSP; P13423; IACC.
 DR InterPro: IPR003896; Binary.toxb.
 DR PRINTS; PRO1391; BINARYTOXINB.
 SQ SEQUENCE 879 AA; 98738 MW; 40685ACB8E05BA01 CRC64;

Query Match 22.6%; Score 886; DB 2; Length 879;
 Best Local Similarity 28.6%; Pred. No. 2e-42;

Matches 270; Conservative 151; Mismatches 270; Indels 252; Gaps 37;

QY 1 MKRKVLPLMALSTILVSSTGNLEVIQAEV-----KOENRLNSESSESSOGILGYFSD 55
 DB 1 MKKKKILGLTC--TVLVGGMTPYPAKITTONTYNDQEVETNEKTVSSGILMGYTFAD 58
 QY 56 LNFQAPNVVTSSTTGDILSPSELENIPISENQ-YFQSAIWSGFIKVKKSDGYTFATSADN 114
 DB 59 EHEKDELIMAPVKNGELFKFNKVEKLTTEKTNIKSIRMTGRILIPSKDGEYTLSTODKN 118
 QY 115 HYTMVDDDOVINKASNSKIRLEKGLYQIKIYOARENTEKGLD-----KLYWTD 167
 DB 119 -VLMQINAE--GEIANTLVNMKGGQYSIRIEIQ-----DDIDGYDDLSPKLYW-E 168
 QY 168 SOKKEVISSDNLQPLKOKSSN-----SRKKRSTSA-GPTVYDRDNDG 211
 DB 169 LNDGKTLIPKKNLFLRYSKSIDENDPRPKDNFEDLKLKRSARLAKGKDEDLDTNDN 228
 QY 212 IPDSLEVEGYTVDKNRKFLSPWISNIHEKGLITKXSSPEKWSASDPYSDEKVTGR 271
 DB 229 IPDAVEKNGTYI---KDSIAVW-EDSFAQGGYKKYLSYLSBSNTAGDPTDYQKASGS 283
 QY 272 IDKNVSEARHPLVAAPYIVHVMENILISKNDSQNTQNDSEKRTSKMTSTST--HT 329
 DB 284 FDAIKAEADPLVAAPYVGVGMEKILISTNEAST---DQKTVSRMTTNSKTDANT 339
 QY 330 SEVHNAEVAH---ANTSTSTSEVHNAEVAHVAIDHSLSLAGERVAETMGLTAD 385
 DB 340 AGVAINAVONGFTGSTITVSHTE-----NSTAVONS---NGE-SMNTSLSTIKGE 388
 QY 386 TARKLNIQVNTGTAIVYVLEPTSLVLGKNOTLATIKAKENQLSQILAPNNYPSKTL 445
 DB 389 SAYINANVRYYNGTAPMYKYVPTTNLVL-DGDTLTITIKADQNGNLSPNETIYPRKGL 447
 QY 446 APTALNAQDDFSTPTMTNNOLELEKTKQLRDTDQVGNATYFENGRRVADT-GS 504
 DB 448 SPLATLMQDFFSSRLIPINDQKLDAGKQIKLETTQVSGN--YGIKNSOGQIILEGN 504
 QY 505 NMESEVLPJOIETARTIIFN-GKDLNLEVERIAVNPSPDETETKRPDMTLEAKIARGFN 563
 DB 505 SMSDYISOIDSLSASITLDGSD--VEFRVYTKADSSNPEDKT-PVLTIGDALEKAGAT 561
 QY 564 EPNGNLQYOGKDLTE--FDENFDQOTSQNTKQNLQELATNATNYTVLDKILNKAMLILR 621
 DB 562 KNGEILYFNGMIDESCVELIFDGNANTANLKERLMAINDKRIYV---QLERKMKLIK 617
 QY 622 D----- 622
 DB 618 TSTFYNNFDGYNFPSSWSNVDSNODGLQMANAKLSGETKIYIPMSKLNPKRYVESGY 677
 QY 623 -----KRFHY-----DRNN----- 631

DB 678 LKNSTSNPIVNTIKAKEQKTYNLVSENDYKFFSEETIGROASNIETITSSGTIFLD 737
 QY 632 -----IAGADESVYKAEHREY-----INSTEGILLN-----I 660
 DB 738 NLSTIELNSPELLEKPIDKVPDSQDEII-DAHKRYVADLSFNQSTANVYLDGLYFEPQT 796
 QY 661 DKDIRKILSGYVIEI-DTEGLKEVINDRYDMLNSSLRODGT-FTDFKKNKPLXI 718
 DB 797 NKEVLDYIQKKEVATLEYSGFKDGTGKDKELRNTYTSNOQKTYNVFNRSY-----FT 850
 QY 719 SNPNY---KVNVAVTKENTINPSENGDSTNGIKILIFS 757
 DB 851 SGENVMPYKKRIYAITPEN-----KELVLVS 877

RESULT 7
 ID Q46221 PRELIMINARY; PRT; 875 AA.
 AC Q46221;

DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE IOTA TOXIN COMPONENT IB PRECURSOR.
 OS Clostridium perfringens.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OX NCBI_TaxID=1502;

QY 1 SEQUENCE FROM N.A.
 RC STRAIN-NCIB 10748;
 RX MEDLINE=94041637; PubMed=8225592;
 RA Perelle S., Gilbert M., Boquet P., Popoff M.R.;
 RT "Characterization of Clostridium perfringens Iota-toxin genes and
 RL Infect. Immun. 61:5147-5156(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NCIB 10748;
 RA Popoff M.R.;
 RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NCIB 10748;
 RA Popoff M.R.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X73562; CAA51960.1; -
 DR HSSP; P13423; IACC.
 DR InterPro: IPR003896; Binary.toxb.
 DR PRINTS; PRO1391; BINARYTOXINB.
 KW Signal.
 FT SIGNAL 34
 FT CHAIN 212
 SQ SEQUENCE 875 AA; 98468 MW; C9AE092CD3818921 CRC64;

Query Match 22.6%; Score 883.5; DB 2; Length 875;
 Best Local Similarity 31.5%; Pred. No. 2.7e-42;
 Matches 262; Conservative 135; Mismatches 279; Indels 155; Gaps 35;

QY 15 TLVSSSTGNLEV-----IQAEVKQENRLNSESSESSOGILGYTSIDLNFQAPNVVTSST 69
 DB 15 TMLTSQTLSTYNVAQTOTNDNTNOKEBITMENTLSSNGLMGYFPADHFDLELMADIKN 74
 QY 70 GDLSPSELENIPISE-NOYFQSAIWSGFIKVKKSDGYTFATSADNHYTMVDDQEVYINK 128
 DB 75 GDLKFEKKYDKLITEDNSSIKSTIRMTGRILIPSDGEYIISTDR-NDVLMQINKAGDIK 133
 QY 129 ASNSKIRLEKGLYQIKIYOAREN-----PTEKGLDFKLYWDSOKKKEVISSDN 179
 DB 134 ---TLKVMKKGAQYINIRIEIODKSLIDMLSYV-----KLYW-ELNGKTVIPEEN 182
 QY 180 LQPLKOKSSNSRKRSTSTAGPTVP-----DADNDGIPDSLEVEG 220

Db	183	LEFRDYKSIDEDND-----PFLPNNNFDFVAFSSAMEDDEDLDTDNDNIPDAYEKNG	233
Qy	221	YTVDYKKNRFTLSPIRISINHHKGLITKSSPEKWSASDPYSDFEYVGTGRIDKNNSPEA	260
Db	234	YTI-----KDSIAVWNKNSFAE-QGYKKVYSYLSNACGPYDYOAKASSIDKAILFA	288
Qy	281	RHPLEAAPIYHVDMDNITLSKNEQDSQOMDSETRFSIKWTSRSPHTSEVGHNAVHA	340
Db	289	RDLPLVAAPVGVGKEMNIIISTNEHASS-----DQKIVYSRAITTSKIDANTV--GVSISA	342
Qy	341	-----NTSTSRTHTSEVGHNAEYHAAVADHSLSLAGEFTMAETMGLTADTARLAN	392
Db	343	GYONGFTGNITTSYSHTTD-----NSTAVQDS-----NGE-SWMTGLSINKGEASAYTAN	391
Qy	393	IRVYVGTGAPRIYNNLPSTSLVIGKNQPTLATIKAEONOLSQLAPANNYPYPSKNLAPIANA	452
Db	392	VRYITGTGAPMKVATPTTNLVL-DGEPLATIKADODNQGNNLSNETYPPKGLSPLALNT	450
Qy	453	QDDESSPTITMNYNOFLELEKTKOLRLDTPDQVYGNIAITYFENGRAVVDTGSNMSEVLPQ	512
Db	451	MDQFARLPIPNYDOLKLDGSKQIKLETTQVSGNYGAKN-SQGOI-ITEGNSMWSNYSQ	508
Qy	513	IQETARLIIFNGKDLNLYERIIAAVNPSPDETETKPPMTLKEALKIPGNEPNNGY-571	
Db	509	IDVSASATILLD-TGSQTFERRVAAKKEQGNPEDKT-PELTIGEALKAPASATK-NGELLYP	565
Qy	572	QGDKITE--PDFNFDOQTQSNIKNOALNATNTYTYLDKIKILAKANNIILRDKRF--HY	627
Db	566	NGIPIDESCVELIFPDNNTSEIIKQGLKLDQKILYV--KLERGNIIILIKVPSYTFNF	621
Qy	628	DR-NNIAVGADESVYKKAHREYVINSSTEGL-----LLNDKDIRKLISGY-----671	
Db	622	DEYNMF--AASMSNIDTQNOGLOSVANKLSGETRIIIPMSKLPYKRYRYSFGSKSPST	679
Qy	672	-----IVEIEDTEGKEVYENDRYMDLNTIS-----SLRODGTFFIDFKYND	712
Db	660	SNSTIVNIAKSKSEQKTDYLVPRKQDTRTKRSYEFETGKSSDLEITLTSGYAFLDNLSTIE	739
Qy	713	-----KLPLYISNPYKV-----NRYAVTKENTIIINSENGDTSNGI	750
Db	740	INSTPEILKEPEIKVPDQELILDHNNKYADIKLDT-----NTGNTYIDGI	785
RESULT	8		
086171		PRELIMINARY; PRT; 721 AA.	
086171			
AC	086171		
DT	01-NOV-1998	(TREMblrel. 08, Created)	
DT	01-NOV-1998	(TREMblrel. 08, Last sequence update)	
DT	01-JUN-2001	(TREMblrel. 17, Last annotation update)	
DE	C2 TOXIN (COMPONENT-II).		
OS	Clostridium botulinum.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;		
OX	Clostridium.		
NCBI_TaxID	1491		
SEQUENCE FROM N.A.			
RC	STRAIN-TYPE C (C)-203028.		
RA	Fujii N., Kubota T., Shirakawa S., Kimura K., Ohishi I., Morishi K.,		
RA	Isogai E., Isogai H.,		
RT	"Characterization of component-I gene of botulinum C2 toxin and PCR		
DR	detection of its gene in clostridial species.",		
EMBL	Biochem. Biophys. Res. Commun. 220:353-359(1996).		
EMBL	D88982; BAA32537.1; -.		

DR HSSP; P13423; IACC.
DR InterPro: IPR003896; Binary toxin.
DR PRINTS: PR01391; BINARYTOXINB.
SO SEQUENCE 721 AA; 80515 MW; 44CB153AC749D5F2 CRC64;

Query Match
Best local similarity 34.5%; Pred No. 1.5e-37.
Matches 221; Conservative 117; Mismatches 221; Indels 82; Gaps 25

20.4% Score 798; DB 2; Length 721;
32 KOENLNESES--SSQGLGYFSDLNFQAQMVYTSSTGDISPSSELENIPEENOYF 89
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 5 KFENEKSNKNKYFTINCLMGYYEFEN-DEFNLNITSPILDGNLTKSKEDINSILG-NKIT 62
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 90 OSATMSGTIXKKKSDYEYFATSAQN-HVTMWDDDEVINKASN-ENKIIFLEGRLYOKI 147
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 63 KSARITIGLKISITGEYLITLSTNSPCRYEL--KEIFLSLNTSYVWLDGNAYIDI 119
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 148 -OYOENEPTEKGLD-FKLWTDSQNKKEVISSDNQLDELKOKSSNRKRSTSAGPTV- 204
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 120 EQLMSENOLLKMYEBIKLYMETSDLIKETIPSEVL---LRPNYSNTNEKSFIPINNLTFL 175
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 205 -----DDRDDGIPLSLLEVGVYVDVYNKKTFFLSPWLSNHHEKKGLKRYKSSPEK.254
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 176 SNAKAKANARNRDPDGDGIDPEMEINGITYVMANKAVAMDCKEPAAN----GYKKIVSNPK 230
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 255 WSTASDPYSDEKYNGRIDKNVSPPEARHPVLAAYPIVHYDMENILSKNEDOSTONTDSE 314
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 221 PCTAMPDPTDEKVSQGOIDPSVSMVARDPMSIAPYIVGQMERLVASKSE---TITGDS 286
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 315 TRTISKNTSTSRTH-----SEVHGAAEV-----HANTSTSRHTSEVHGAAEV 358
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 287 TKSMKSTSHSTNTINTVGAEVSGSLQLAGCIPVEFSMASANSYHMTQNTSTVD---- 341
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 359 HAVAIDHSLAGEPYTAETMGLNADPARLANRYVTCGAPIYNYLPPTSILYGKNQ 418
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 342 -----DTTGE-SFGQGLSTINGESAYINPRIYNTGTAPAYNYNPPTTYIYDK-Q 390
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 419 TLATIRAKENOLSQLAPANNYPYPSKNLAPIALMAODDFSSPTITMANYNOFELEKTKOLR 478
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 391 SVATIKGEESLIGDYLNFGCTYPIIGEPPMALMTMDQESSRIPIPTYNOLKSIDMGCTVA 450
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 479 LDTPDVVNINATPVNENGRVRNDGSNMNSEVLPLOETFARLI--FNCGDLNVERRIA 536
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 451 LSTSOFTENFAKYN-SNGNLVTD-GNMNGPYICTIKSTASTLASLSPSGTTOVA---VYA 505
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 537 VNPSPLETTPDMTLKEALKIAGFPNPGNLIYOQKDIT---EFFDNFOOTSONIKN 593
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 506 PNFSPEDEKT-PKLTLEQALVKAFALKEKKGKFYFHGLEISKNEKIQVELDNTNNDPEN 564
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 594 QLAELMATNITYVDKIKLNKMKMLLIDKRRHYRNRTAV 634
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 565 QLKTNADKDIMHC1-IKRN-MNILVKVITFKENISSINI 601
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

RESULT 9 PRELIMINARY; PRT; 204 AA.

ID O9X377
AC O9X377;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PX01-111.
OS Bacillus anthracis.
OC Plasmid virulence plasmid PX01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxId=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=STERNE;
RA Okinaka R.T., Cloud K., Hamton O., Hoffmaster A., Hill K.K., Keim P., Koehler T., Lamke G., Kumano S., Mahillon J., Manter D., Martinez Y.,

RA Rieke D.O., Svensson R., Jackson P.J.;
RT "The sequence and organization of pXOI, the large Bacillus anthracis
plasmid harboring the Anthrax toxin genes.";
RL J. Bacteriol. 0:0-0(1999).
DR EMBL: AF065404; AAD32415.1; -
DR HSSP: P13423; IACC.
KW Plasmid.
SQ SEQUENCE 204 AA; 23029 MW; E1657B23AE4273FD CRC64;

Query Match 6.0%; Score 235.5; DB 2; Length 204;
Best Local Similarity 34.6%; Pred. No. 2,7e-06;
Matches 56; Conservative 36; Mismatches 47; Indels 23; Gaps 6;

QY 616 MNLIRDRKRRHYDRNNANVGADESVKAEHREYNSSTEGLLINDIDIRKILSGYIEI 675
D 1 MNLIVRDP-YHYDNGNIVGVDSYLKNAKQILNMSSDGSLLDLDDVNALSGYMIQI 59
QY 676 EDTE-----GLKEVINRDMINISLRDQKTFIDFKRYNDKLPYISNP 722
D 60 KKSNNHILNPSVITTLGKRSQGVGLRYVLS-----DGIQFLDPKFKDEMWSLV-DPG 112
QY 723 YKVVNVAVTKEK-TIIPSENGDTSTNGIKKILFSKKGYEI 763
D 113 DQVYVAVTKEDFNVAVTRDENGNTA-NKLNKTLVLSGKIREI 153

RESULT 10
Q9PQ08 PRELIMINARY; PRT; 4688 AA.
ID Q9PQ08;
AC Q9PQ08;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE HYPOTHEICAL PROTEIN U0482.
GN U0482.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID-134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROVAR 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Letkovitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
urealyticum.";
RL Nature 407:757-762(2000).
DR EMBL: AEO02145; AAF30894.1; -
DR InterPro: IPR001152; Thymosin_b4.
DR InterPro: IPR000626; Ubiquitin.
DR Pfam: PF00240; ubiquitin.1.
DR SMART: SM00152; TRY.1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 4688 AA; 534880 MW; B53ABFAFEE1997E CRC64;

Query Match 5.3%; Score 205.5; DB 2; Length 4688;
Best Local Similarity 20.8%; Pred. No. 0.011;
Matches 176; Conservative 138; Mismatches 337; Indels 197; Gaps 42;

QY 18 VSTGTGLEVI-QAEVKOENRLNSESOGGLGYFSDLNFOAPMVVYTSSTGDLSTPS 76
D 3676 INTGKGEVLFVSKGLKSNOLYK-----LVVVYILD-NIHONIDETKRIFFSDHV-S 3725
QY 77 SELEINISENQYFOSATISGFIKVKSDGYTFATPSADNHYTMVDDDEVINKASNSKIR 136
D 3726 KELEINQVIMISKHGMKSPDTTANFERKIEFYQ-----DDNDVNLNDAIVKFK 3776
QY 137 LEKRLQYIKIORYORENPTEKGLDKLYMTDSQNKKEVSSDNLQLELKQSSN---S 192
D 3777 DEHNINIKQIKIVRIKEN-----NDMLIKQI---DNLN-PETKKLENIELSK 3820

QY 193 RKRSTSGAPVPPDR-----NDGIPDSLEVEGYVDVKKKRTFLS 233
D 3821 PLKTHNLNLSVINDKENSLITTEGNPLKAYIQONDTINTQOTINTVLSGVSK-YNG 3879
QY 234 PWISNHEKKGLTRYKSSPEKWSASDPYSDFEKVTGRIDKNVSPKARHPLVAAYPIVH 293
D 3880 RQIKVYVKKDNNNVIVES---LITLQKGKNDYQLLSNLNSN-----REYFEKEIETHI 3931
QY 294 -----DME-----NIIISKNEQSTQNTDSE-----TETISKNTSTHSEVGN 335
D 3932 SNFNNEFDELEKNGVSNFIQTQKNTTYQNMDSATITGVTFGVNFNFKI-KSEDKILENN 3990
QY 336 AEV---HANTSTSRTHSEVGNAEVNAHVAIDHSLSLAGEPTMETGLNT---ADARL 389
D 3991 QQVVAWEPKPEKTIIDTNTWLOYTRPLKDVISDEK-----EGTMAHDSNSNFEKEETYYK 4045
QY 390 NANIRVYVTGAPLYNVLPPTSLVGLKNOQLA-----TIKAKENQSLQILAPNNVYPSK 443
D 4046 LKIQFVKKPFKAKNNINNSNNVILDTNTNSINSNYEFTTKVGDKLINTSSNNVNTNS 4105
QY 444 NLADIALN-AODPSSPEITVNY---NOFLEKTKQLRDTDOYVGNATVNFENG-- 497
D 4106 QTIHFETLSGVKKSWMGKKIKLSYKSDNTSESIHTNEVLEESNKTQYNTILNKLKRNRTYT 4165
QY 498 ---VRVDGSMWSEVLPOIQTETARIIFNGKDLNVERIAVNPSPDLETT-----K 547
D 4166 LIDVKKLIDNNVSDPPEKGNLTNSPTTTSAINVLNIEISNASTNPKSTIIKININD 4225
QY 548 PDMTLKEA-----LKIAGFEPNPNLOYOGKDTTEPFNFDOQTSQNIK-NOLAE 597
D 4226 PDNVLRDQDQATYVGNKKQAMGFITVSGNIKYLATIVDLNPN-DKYNIVNISFNKPS 4284
QY 598 LNAITNITVLDKIKLNKAKNMLINDKRFHYD-----RNNIAYGA--DESVYKAEHRE 647
D 4285 IAAEN-----IGIDKSNNTI-----YNDISIPKLEINNDIIVGPIKEIYVKNAMOK 4332
QY 648 VINSTEGLLINDKDIRKILSGYIELEDFEGLEKEVIN--RVDMLISL-RQDGTF 704
D 4333 --NNIDVGLQIINKIAHNLR-FLAKFKSTN-----NDIIEIETVINGSSLVVNDGTF 4383
QY 705 IDF-----KKY-----NDKLPYISNPYKVVAVTKEHTIIN 738
D 4384 IRLFLNNKAKKILSVLVVYLVVNNNSNTVYESNKL-P-KLNNINVIQK---INSHITII 4438
QY 739 PSENGDTS 746
D 4439 -SKNGEMS 4445

RESULT 11
Q9GZ76 PRELIMINARY; PRT; 1387 AA.
ID Q9GZ76;
AC Q9GZ76;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, last annotation update)
DE REFICULOCYTE-BINDING PROTEIN 2 HOMOLOG A (FRAGMENT).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID-5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20402589; PubMed=10920203;
RA Rayner J.C., Galinski M.R., Ingravallo P., Barnwell J.W.;
RT "Two Plasmodium falciparum genes express merocrite proteins that are
related to Plasmodium vivax and Plasmodium yoelii adhesive proteins
involved in host cell selection and invasion.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:9648-9653(2000).
DR EMBL: AF196347; AAF98066.1; -
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 1387 AA; 164871 MW; 5660154744F0C285 CRC64;

Db 1708 MDGPISNVNPEELAHVAEGSKLEAKERSMDADK---TTEEDIT-----VEDPBG 1758
QY 675 IEDTBELKEVINDRYDMLNSSLROGKPTIDKKYNDKPLIISNPYVNVYAATKEN 734
Db 1759 IGHQNLKEVHEQASBELNTYNSL--DGRINVEKRELDPGSIPIIDRITTEHIEIDREK 1816
QY 735 TIINPSE---NGD 744
Db 1817 EIHEPNELOAHNGE 1830
RESULT 13
Q9BK45
ID 09BK45 PRELIMINARY; PRT; 3254 AA.
AC 09BK45
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE RETICULOCYTE BINDING PROTEIN 2 HOMOLOG B.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21101060; PubMed=11160005;
RA Triglia T., Thompson J., Caruana S.R., Delorenzi M., Speed T.,
Comman A.F.;
RT Identification of Proteins from Plasmodium falciparum That Are
RT Homologous to Reticulocyte Binding Proteins in Plasmodium vivax.";
RL Infect. Immun. 69:1084-1092(2001).
DR EMBL: AF312917; AAK19245.1; -
SQ SEQUENCE 3254 AA; 382876 MW; 6F9CAFA5AA6167BA CRC64;

Query Match 5.0%; Score 195.5; DB 5; Length 3254;
Best Local Similarity 19.2%; Pred. No. 0.024;
Matches 172; Conservative 140; Mismatches 294; Indels 291; Gaps 41;

QY 5 KVLIPMALSTIIVSSGTGNEVIAEQAEVKQENRLNSESSESSGGLGYFSDLNFOAPMV 64
Db 1920 KHVIKLANFSGIIVMSDTNTEITPENPLENDLLN-----LQLYFERKHEI 1965
QY 65 TSTTGDLSPSSELENI--PSENOYFQSAIWSGFIVKKSDEYTPATSDNHTMVVD 122
Db 1966 TSTLENDSDL--ELDHLSNSDESIDNLKVYNDIEL-----HYST-----QILKYLDN 2013
QY 123 -----OEVIKASN--SNKIRLEKGRLYOI-KIQ 148
Db 2014 IOKLKGDCNDLVKCKELRELSTALYDLKIQITSVINRENDISNNIDIVSNKLINEIDAIO 2073
QY 149 YORE-----NPTKGLDFKLYMTDSQNKKE 173
Db 2074 YNEEKKEIFDNEVEYKTLDDTKNAVYVKAELIKVNDINKTKEDLDIYFNDDLEKSL 2133
QY 174 VISSDNLQJPELKQSSNRKRSSTAGPTVPDRDNG--IP--DSLEVEGYTVD--- 224
Db 2134 TSSNMEIKTIYQNSYNS--FSDINKNINDIKEMKTLIMLDELLEGNINIDISLY 2189
QY 225 ---VKNKRTFLSPWISNIHEKGLT---KYSSPEKMWSTAS---DPSDFEYKVRID 273
Db 2190 NFIRNIQIKIGNDIKINREQEDWNTICEFYIGNNNYFIKSDISIFNKYDDHKKVONYIS 2249
QY 274 KAVSPAPRPLVAAPIYIVVD--MENI---ILSKNED-----QSTQ-----NTDSE 314
Db 2250 NNIDVYVKNKNSLSSEHVINATNIENIMTSIYINEDTEKNSLEETQDKLLELYENFKKE 2309
QY 315 TRTISKSTSTR--TRTSEVHGAENVAHANTSTRHTSEVHG-----NAEV 358
Db 2310 KNINNNYIVHFNKLEIENSLLEYINSISTNKNKINETONIDILKNFNNITKTKNDKY 2369
QY 359 -HAVAIDHLSLAGERTMAETMG--LNTADTARL---NANIRYVNTGTAPYVNLPTPT 410
Db 2370 KELVHVDSTLTLESIGTFNNLVGDLMSNIQDQVYKYEDINNVELKAKVLYIENTNTLIGRI 2429

QY 411 SLVIGKNOITLATIKAKENOIS---QILAPNNYPSKVLAPIALNAODESPPTIMYNO 467
Db 2430 NTEI---KELDKYQDENNGGIDKXYIEINKENNSYIIR-LKEKANLKNFIS-----K 2476
QY 468 FLELEKTKQRLRDLDDQYVGNIAATYFENGVRVDTGSSNMSEVLPIQIDETARIIFNCKDL 527
Db 2477 LLONIKRNREL-----YNINN--IKRD-----IMNTGKSV 2505
QY 528 NLVERRIAAYNPSPDLETTKPDMLTKALKIAGFENPENGNOYQGKDITFDFNFDOOT 587
Db 2506 NNIKOKFSS-----NLPLKREL--FQMEMLINI----- 2532
QY 588 SONIKNOJLAELNNTIYTVL-----DKIKLAKNMILLROKRF-HYDRNNIATGADES 639
Db 2533 -NNIMNETKRISMTDATYNTTLQDIENNNKKNENNMNIETIDKIDIKIKHNEKIQAEIL 2591
QY 640 VKEAHREVINSSTEGILLNIDKIRKILSGYIYEIDTEGLKEVINDRYDMLNSSLRQ 699
Db 2592 IIDDARKKV-----KETTDINKAFNEITEYNNNE-----NNGVIR 2627
QY 700 DGKTFIDFKRY-NDKLPYIISNPYKVYVAVTKENTINPSENGDSTNGIKKILI 755
Db 2628 SAKNIYDVKATYLLNNELDKFL-----LKNELLS-HUNNDIKDGDDEKLIL 2671

RESULT 14
Q9BK46
ID 09BK46 PRELIMINARY; PRT; 3130 AA.
AC 09BK46
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE RETICULOCYTE BINDING PROTEIN 2 HOMOLOG A.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21101060; PubMed=11160005;
RA Triglia T., Thompson J., Caruana S.R., Delorenzi M., Speed T.,
Comman A.F.;
RT Identification of Proteins from Plasmodium falciparum That Are
RT Homologous to Reticulocyte Binding Proteins in Plasmodium vivax.";
RL Infect. Immun. 69:1084-1092(2001).
DR EMBL: AF312916; AAK19244.1; -
SQ SEQUENCE 3130 AA; 370415 MW; 13D973DB89D82026 CRC64;

Query Match 4.9%; Score 193.5; DB 5; Length 3130;
Best Local Similarity 19.7%; Pred. No. 0.03;
Matches 173; Conservative 142; Mismatches 310; Indels 253; Gaps 44;

QY 5 KVLIPMALSTIIVSSGTGNEVIAEQAEVKQENRLNSESSESSGGLGYFSDLNFOAPMV 64
Db 1920 KHVIKLANFSGIIVMSDTNTEITPENPLENDLLN-----LQLYFERKHEI 1965
QY 65 TSTTGDLSPSSELENI--PSENOYFQSAIWSGFIVKKSDEYTPATSDNHTMVVD 122
Db 1966 TSTLENDSDL--ELDHLSNSDESIDNLKVYNDIEL-----HYST-----QILKYLDN 2013
QY 123 -----OEVIKASN--SNKIRLEKGRLYOI-KIQ 148
Db 2014 IOKLKGDCNDLVKCKELRELSTALYDLKIQITSVINRENDISNNIDIVSNKLINEIDAIO 2073
QY 149 YORE-----NPTKGLDFKLYMTDSQNKKE 173
Db 2074 YNEEKKEIFDNEVEYKTLDDTKNAVYVKAELIKVNDINKTKEDLDIYFNDDLEKSL 2133
QY 174 VISSDNLQJPELKQSSNRKRSSTAGPTVPDRDNG--IP--DSLEVEGYTVD--- 224
Db 2134 TSSNMEIKTIYQNSYNS--FSDINKNINDIKEMKTLIMLDELLEGNINIDISLY 2189

```

QY 225 ---VKNRFTLSPWISNHEKKGLT-----KXSSPEKSTAS-----DPYSPFEKYTGRIID 273
Db 2190 NFIIRNIOIKIGNDIKNREQENDPTICEFYIQNNYNEIKKSDISIFNKDYDHIKVNDYS 2249
QY 274 KNSPEARHPLVAAPYIVHV--MENI---ILSKNE--OSTONTSETPTISKNTSTS 325
Db 2250 NNIDVKNHNSLSEHYINAINIENITMTSIVEINEDPTSMLEETQOKLEIYENFKE 2309
QY 326 RTHSEVHNAEVAHANTSTSTHTSEVHNAEVAHAVALDHSLSLAGEPTMAETGMLTAD 385
Db 2310 K---NIINNNKI-----VHFN---KLKEIENSL-----ETYN----- 2336
QY 386 TARLANANRYVNTGAPRYNVLPPTSLVGLGNQTLATIKAEQ-----LSQ 432
Db 2337 --SISTENKIN-----ETONIDILKNE--FNNIKTRINKVKELVHVDSTLTLES 2383
QY 433 ILAANNYPKSLAPIALNADDESPRTMNYNOFLEKTKOLRLDPTQVYGNATYN 492
Db 2384 IQTFNNLY-----GDMNSNIODYKYKEDINN-----VELKKVKLYIENTITLGRINTFI 2433
QY 493 FENGRRVDTGSNNSEVLPOIOETTARII--FNGKDLNIVERRIAANPSDPLETTKPDMT 551
Db 2434 KELDKYQ--DENNGIDKYEIKENNSYIILKEKANNLKE-----NFSKLLQNIKRNET 2486
QY 552 LKEALKIAGFNEPENGNOYOGKIDTEDEFNPDQTSQ--IKNQLAE-----LNATYIY 604
Db 2487 -----ELYNNINIKRDIIMTGSVYN-----NIKQFSSNLDPLKEKLFQEMMLNINIM 2536
QY 605 TVLDKIKLNAA--MNLIRD--KRFHYDRNNIYAGADESVKAEHREYVNSSTEGILLNID 661
Db 2537 NETRISTATVNTITLQDIENNNKKNENNNIETIDKID--HIKINKEIQAEILID 2594
QY 662 KDIRKILSGYIEIEDTEGLEKVIN--DRYDMLNSSLRODGKTFIDFKKY--NDKPLV 717
Db 2595 DAKRRVK-----EIDNINIKAFNETEYNNYNNEN--NGVKSANNIYDEATYLNNEIDKF 2646
QY 718 ISNPNKYVNAVYKENTIIINPSENGDSTNGIKKILI 755
Db 2647 L-----LKNLELLS--HNNNDIKLDGDEKLIL 2671

RESULT 15
09ZHL0
ID 09ZHL0 PRELIMINARY: PRT: 4919 AA.
AC 09ZHL0;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE LARGE SUPERNATANT PROTEIN 2.
GN LSPA2.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_Taxid=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=35000.
RX MEDLINE=99030326; PubMed=9811662;
RA Ward C.K., Lumbley S.R., Latimer J.L., Cope L.D., Hansen E.J.;
RT "Haemophilus ducreyi secretes a filamentous hemagglutinin-like
protein."
RL J. Bacteriol. 180:6013-6022(1998).
DR EMBL; AF057696; AAC79761.1;
DR InterPro; IPR00130; 2n_MTPeptide.
DR SMART; SM00235; znmC; 1.
SQ SEQUENCE 4919 AA; 542602 MW; 5779201455CA69A0 CRC64;

```

Query Match 4.98; Score 193.5; DB 2; Length 4919;
 Best Local Similarity 21.18; Pred. No. 0.057;
 Matches 180; Conservative 131; Mismatches 324; Indels 219; Gaps 42;

QY 1 MKKKRVILPLMALSTILVSTGNLEVIQAFKQENRLNSESSESSQGLGYFSDLNFOA 60

```

Db 245 VQRGVILIGKDVAF--NGLSHPDVAKNIEQQKVISIEDSPAKL-----ANVFAA 296
QY 61 PMVVTSTTGDLSIPSELENIPESENOYFOGAINSG-----FIKKKSEDTYF 108
Db 297 GNLYDVNTROVNNNTNKKRPITDTRKDNIAI--SGEAGSMYGRNIKFYTDG----- 350
QY 109 ATSAADNH--VTMWDDQEVINKASNSKRILEKRLYQIKIQYQRENTEKGLDFKLYWT 166
Db 351 --AGVNHQGVIFAEDDILNITDDGNS-----RLNKVYADYVR-----VVGKDIEL-- 393
QY 167 DSQNKKEVISSDNLQPLBELKSSNSRKRKTSAGPYTPDDNDGICPD--SLEVEGYVDV 225
Db 394 -----ANNOIHAQDQOLINAGHVKLVNGSSVISNNNGISALNITLENATVSA 443
QY 226 KNRFTLSPWISNHEKKGLTKRYKSSPEKSTASDPYSD-----FEKYTGRIIDN 275
Db 444 NN-----LSFRYNTDKLNNLSKVS-----ARADLQSGNLNDKASVLAHKLTLNLSND 493
QY 276 VSPKARHPLVAAPYIVHVDENIILSKNEOSTONTSETPTISKNTS--TSRTHSEV 332
Db 494 VSLNNSQSLSA-----NNLKIKKVRDLNLSSELANNLTLNNSNITLKNKSKFT 544
QY 333 HGNAEVAHANTSTSTHTSEVHNAEVAHAVALDHSLSLAGEPTMAETGMLTADTARLAN 392
Db 545 AGNMTLANTNVNTLNDSLANLTLNVTKNVTLNDASKS--ANKLDLNTDVTNLSK 603
QY 393 IRVYNTGAPRYNVLPPTSLVGLGNQTLATIKAEQSLQILAPNNYPSKNLAPIALNA 452
Db 604 -STLSAGLTFKKV-----KNVTL-----NNNSE--LAANN-----LSLNA 636
QY 453 QDDFSSPTITMNYNOFLEKT--KQLRL--DTPQVYGNATYTNFENGRRVDTGSWMS 507
Db 637 -----SHVVTLNNSKLSAQAKADIKAVNLTLNDTE-----LTANKLDINSTYIINGTIA 687
QY 508 EVLPQIOETTARIIFNGKDLNVERRI--AAVNPSPDLETPDPTLKEALKIAGFNEPN 566
Db 688 GIFANI--TTEKLNKKEKALLAEONLNFYNGSH--YENKGDIVSKDAVTFSKNS-- 741
QY 567 GNLOYOGKIDTEFPNPDQTSQNIKNQLAEIENATNI-----YTVLDKIKLNAAK----- 615
Db 742 -----DPTSGSKLVNAQNL--KVNNNFTISOGDITTLIGNTLVNLSGFT 787
QY 616 -----MNLIDKRRHYDRNNIYAGAD--ESYVA-----EAREVI 649
Db 788 NSGNLTIVTKLIDVGDIDQFTNGMLTVGEDLHISKTKITNDGKLSIKNLSSEADFI 847
QY 650 NS-----STEGLLNIDKDIRKILSGYIEIEDTEGLEKVINDRYDM--LNIS 695
Db 848 NNGTILGTEALKINTKGNFTKKEKAI--LASNSLDISVACGKTFNNGTIESGKNLNT 905
QY 696 S-----LRQDGKTFIDFKKYNDKPLIYSN-----PNKVVNAVYATKENTIIINPSENGDTS 746
Db 906 NTGAFLVANDNATIRSFGLNITSTGNSNNGTLISNERLN--ITSANFTNES--NGTYM 961
QY 747 TNGIKKILIFSKG 760
Db 962 SNGLLNIT--AKQG 973

```

Search completed: December 2, 2001, 13:54:52
 Job time: 527 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:51:49 ; Search time 44.8 Seconds
(without alignments)
625.266 Million cell updates/sec

Title: US-09-747-521-4
Perfect score: 3913
Sequence: 1 MKRRKVLIPALSTILVSS.....TSTNGIKKILFSKKGVEIG 764

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3800	97.1	764	1 PAG_BACAN	P13423 bacillus an
2	235.5	6.0	192	1 YPA_BACAN	P13422 bacillus an
3	183	4.7	1803	1 YJL3_YEAST	P47024 saccharomyc
4	182	4.7	1658	1 YMG7_YEAST	003661 saccharomyc
5	173.5	4.4	1230	1 SMC3_YEAST	P47037 saccharomyc
6	173.5	4.4	1271	1 Y338_MYCGE	P47380 mycoplasma
7	166.5	4.3	1928	1 MYSL_YEAST	P08964 saccharomyc
8	162	4.1	2116	1 MYSL_YEAST	P08964 saccharomyc
9	162	4.1	2869	1 RBPI_PLAVB	000798 dictyostell
10	161	4.1	1957	1 YD86_SCHPO	010411 schizosacch
11	160	4.1	1790	1 US01_YEAST	P25386 saccharomyc
12	159.5	4.1	1630	1 MSPI_PLAFK	P04932 plasmodium
13	159.5	4.1	1639	1 MSPI_PLAFK	P04932 plasmodium
14	158.5	4.1	1487	1 MDS2_YEAST	P53094 saccharomyc
15	158	4.0	1385	1 PART_SCHPO	013735 schizosacch
16	157	4.0	1000	1 S155_YEAST	P43612 saccharomyc
17	156.5	4.0	803	1 SM16_YEAST	P09959 saccharomyc
18	156.5	4.0	1208	1 PCPI_SCHPO	Q92351 schizosacch
19	156.5	4.0	1251	1 RBPI_PLAVB	000799 plasmodium
20	156	4.0	1744	1 TANX_XENLA	001550 xenopus lae
21	155.5	4.0	1018	1 FNBA_STRAU	P14738 staphylococ
22	155.5	4.0	1024	1 R1P3_MOUSE	P97434 mus musculu
23	155.5	4.0	1358	1 S1R4_YEAST	P11978 saccharomyc
24	155.5	4.0	1541	1 IGAI_HAEIN	P42782 haemophilus
25	155.5	4.0	1545	1 IGAI_HAEIN	P45385 haemophilus
26	153	3.9	1570	1 P3K1_DICDI	P54673 dictyostell
27	153	3.9	1726	1 MSPI_PLAFK	P50495 plasmodium
28	151.5	3.9	2334	1 WAPA_BACSU	007893 bacillus su
29	151	3.9	1218	1 MGPC_MYCPN	Q50341 mycoplasma
30	150.5	3.8	730	1 GLN3_YEAST	P18494 saccharomyc
31	150.5	3.8	1116	1 YK54_AQUAE	O67838 aquifex aeo
32	150	3.8	1420	1 SRB9_YEAST	P38931 saccharomyc
33	150	3.8	3418	1 BRCA2_HUMAN	P51567 homo sapien

ALIGNMENTS

RESULT	1	STANDARD	PRT	764 AA.
PAG_BACAN				
AC	P13423;			
DT	01-JAN-1990 (Rel. 13, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	PROTECTIVE ANTIGEN PRECURSOR (PA) [CONTAINS: PA-20; PA-63].			
GN	PAG.			
OS	Bacillus anthracis.			
OG	Plasmid pXOI.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/staphylococcus group; Bacillus.			
OX	NCBI_Taxid=1392;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MDLINE=69172073; PubMed=3148491;			
RA	Welkos S.L., Lowe J.R., Eden-McCutchan F., Vodkin M., Leppla S.H.,			
RA	Schmidt J.T.;			
RT	"Sequence and analysis of the DNA encoding protective antigen of			
RT	Bacillus anthracis.";			
RL	Gene 69:287-300(1988).			
RN	[2]			
RP	DOMAINS.			
RX	MDLINE=91332080; PubMed=1651334;			
RA	Singh Y., Kimpel K.R., Quinn C.P., Chaudhary V.K., Leppla S.H.;			
RT	"The carboxyl-terminal end of protective antigen is required for			
RT	receptor binding and anthrax toxin activity.";			
RL	J. Biol. Chem. 266:15493-15497(1991).			
RN	[3]			
RP	X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).			
RX	MDLINE=97192099; PubMed=9039918;			
RA	Petosa C., Collier R.J., Kimpel K.R., Leppla S.H., Liddington R.C.;			
RT	"Crystal structure of the anthrax toxin protective antigen.";			
RL	Nature 385:833-838(1997).			
CC	-1- FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,			
CC	AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE			
CC	DEATH. PA IS THOUGHT TO BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC			
CC	CELLS, THEREBY FACILITATING THE INTERNALIZATION OF LF OR EF. PA			
CC	ASSOCIATED WITH LF CAUSES DEATH WHEN INJECTED, PA ASSOCIATED WITH			
CC	EF PRODUCES EDEMA. PA INDUCES IMMUNITY TO INJECTION WITH ANTHRAX.			
CC	-1- SUBUNIT: ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT PROTEINS, A			
CC	PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN EDEMA FACTOR			
CC	(EF). NONE OF THESE IS TOXIC BY ITSELF. ONCE ACTIVATED, PA FORMS			
CC	HEPTAMERS WHICH INSERT INTO MEMBRANES AND FORM CATION-SELECTIVE			
CC	CHANNELS.			
CC	-1- SUBCELLULAR LOCATION: SECRETED.			
CC	-1- DOMAIN: THE C-TERMINAL PART OF PA IS REQUIRED FOR RECEPTOR BINDING			
CC	AND TOXIC ACTIVITY.			
CC	-1- PTM: PROTEOLYTIC ACTIVATION BY FURIN CLEAVES THE PROTEIN INTO TWO			
CC	PARTS, PA-20 AND PA-63, THE LATER HEPTAMERIZE.			
CC	-1- SIMILARITY: TO C. PERFRINGENS IOTA-B TOXIN AND TO VIP1 TOXINS IN			
CC	BACILLUS.			
CC	*****			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			

CC Saccharomycetales: Saccharomycetaceae: Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / FY1679;
 RX MEDLINE=97103775; PubMed=8948101;
 RA Czaplich C., Korde E., Pujol A., Jauniaux J.-C.;
 RT "Sequencing analysis of a 4.0 kb fragment of yeast chromosome X
 RT reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,
 RT SPI10, GCD14, REP1, PHO86, NCA3, ASF1, CCT7, GEF3, two tRNA genes,
 RT three remnant delta elements and a Ty4 transposon.";
 RL Yeast 12:1471-1474(1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 249389; CAAB9409.1; -
 DR SGD: S0003649; YOL113W.
 DR InterPro: IPR001584; Rve.
 DR InterPro: IPR001878; znf_CCNC.
 DR Pfam: PF00665; rve; 1.
 DR SMART: SM00343; znf_C2HC; 1.
 KW Transposable element; Hypothetical protein.
 SQ SEQUENCE 1803 AA; 207693 MW; 16DCD7284A8D52D3 CRC64;

Query Match 4.7%; Score 183; DB 1; Length 1803;
 Best Local Similarity 20.1%; Pred. No. 0.14;
 Matches 167; Conservative 117; Mismatches 283; Indels 264; Gaps 40;

OY 25 EVIOAEVKOEENRLNSESOGILGYRPSDLNFOAPMVYVTSSTGDLSPSSLELNIPS 84
 DB 1073 ELIQEBLKKTN---HETSPKESIG---TNKFRNTNNEISLKTDDTSIPKLTLSINN 1126
 OY 85 ENQYFOSAIWSGFIKVKSDERTFATSADNHVTWWDDEVINKASNSKIRLEKGRLYO 144
 DB 1127 HH-----SNDY-----STNKVE----- 1138
 OY 145 IKIOYOREN---PTEKGIQFKLYMTDSQKKKEVISDNQLPELKOKSSNKKRSTSA 200
 DB 1139 ---KKEKENHHPREDIDYDM---SDQDMESNCQDGNLKLKELVTDKNVPTDNGTNY 1190
 OY 201 GPTVPDRNDGIPDLSLEVGTYVDVKNKRTFLSPWISNIHEKK-----GLTRY--- 248
 DB 1191 SPR-----LEQNEASGSPYQYVKNKSAFLNKEFSSSLMKRRKRHRDKNNLSLTSELE 1242
 OY 249 -----KSSPEKWSSTADSPYSDFEKYVIGRIDKNVSPEARHPLVAAYPIVHVDMEN 297
 DB 1243 RDKRKRKKNRVKLIPDNMETVSAPKIRAIYVNEAISKNEDLKEKHVKQAY---HKEION 1299
 OY 298 ILSKKEQDSTQNTDSEETISKN--TTSRRTSEVHG--NAEVANANTSTSKTHSEVH 353
 DB 1300 LQDMKVFVDVAVYSRSE---IPDNLIVPNTJFTKRRNGIKYKARIYCRDQTSPPDYTVI 1356
 OY 354 GNAEVAIVAVIIDSLSLAGERT-WAETMGLNTADT-ARLANAIRYVVTGTAIVNVP-RT 410
 DB 1357 TTESLNHNHIKIFLMTANRRNFMKLTLDINHAFLYAKLEE-----IYIPRHR 1406
 OY 411 SLVLGNQTLATIKAKENO-----LSQI-LAPNNYVP-----SKNLAPIALNAOD 455
 DB 1407 RCVVVKLNKALYGLKQSPKEMNDHLROYLNGIGLKDNSYFGLYOTEDKNLM-LAVYVDDC 1465
 OY 456 FSSTPTMYNOFL-ELEKTKOLR-----LQTDQYGNATATNENGGRV----- 498
 DB 1466 VLAASNEQRLDEIFINKLSNFELKITGTLIDVLDLTD-LIGMDLVYNNKRGITDLTKSP 1524
 OY 499 --RVDTGSNMSEVLPQIOETTARIIFNGKDLNVERRIAFAVNP-SPLEETTKPD-----M 550
 DB 1525 INRMD--KKRYNEBLKIRKSSIPHMSTYK-----IDPKDVLQWSEEFROGYL 1571
 OY 551 TLKEALKIAGFNEPNCNLOYGKDDITEFPNPDQOSTQINIKQALAEANTNTYTLDKI 610
 DB 1572 KIQOOL-----GELYV--RHRCRYDIEFAVVKVRLVYPEREFYMTYKTIQYL 1620
 OY 611 KLNAMKMLIRDRFPHDRN-----NIAGDASVYKEARE-----VI----- 649
 DB 1621 -----VRKDDGIHRDRCNNDKVIATIDASVGSSEYDAQSRIGVITLGMNINPVIS 1673
 OY 650 NSSTEGILLNIDKDIRKILSGY-----IWEIEDT-----EGIK-- 682
 DB 1674 NKSTNCVSTSEAEHLAIYEGYADSETLKYTLKELGQNDNDIVMTDSKPAIOGLNRSY 1733
 OY 683 -----EVINDRYMLNLISSLRQDGKFI-----DFKKY 710
 DB 1734 QQPKERFTWIKETIIEKIKESIKILKTIGKNIDLLTRPVASDPKRF 1784

RESULT 4
 YM67_YEAST
 ID YM67_YEAST STANDARD; PRT; 1658 AA.
 AC 003661; Q04988;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHETICAL 187.1 KDA PROTEIN IN GUAL-ERG8 INTERGENIC REGION.
 GN YMR219W OR YMR261.13 OR YMR959.01.
 OS Saccharomycetes cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE OF 1-711 FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
 RL Walsh S.V.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 608-1648 FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A.,
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 249809; CAAB9934.1; -
 DR EMBL: 249939; CAAB9190.1; -
 DR SGD: S0004832; YMR219W.
 KW Hypothetical protein.
 SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;

Query Match 4.7%; Score 182; DB 1; Length 1658;
 Best Local Similarity 19.3%; Pred. No. 0.14;
 Matches 161; Conservative 133; Mismatches 308; Indels 234; Gaps 39;

OY 25 EVIOAEVKOEENRLNESE-SSSOGILGYRPSDLNFOAPMVYVTSSTGDLSPSSLELNIP 83
 DB 287 EGYEMLEDDIDVSEKDEKESQGAECTEHS-VDFSXYWQPRDNTKIPVTEKYESD 345
 OY 84 SENQYFOSAIWS-GFIKVKASDE-----YTFATSADN---HYTMVVDQEVYINKASN 131
 DB 346 VHQYSEDAQAFDGSVNIYVDSDESEDESAESYSANAENVYHNEHLDKDELTIEDIES 405
 OY 132 SNKIRLEKGLYQIKIOYORENPTKGLDKL-----YWTDSQNK 171

Db 406 SDS-----ESQSAQESQGSIDPEFYKMKNEKSTSETEUNTSESROGPAKDAVTK 456
Qy 172 KEVSSDNLQIPEL-----KOKSSNRKRRKSTAGPTVPDRDNDGIPDSL 216
Db 457 NKVEQENDEPEREKDILRSIDKNFHCNNKSEYSEVNLNETDPAIVERNO-IND-- 513
Qy 217 EVEGYTVQVKKRRFL---SPWISNIH- KKGLTKYSS-----PEKWTASDPY---S 263
Db 514 -VEGYDVGKSVESDLHESP--DNLXDLAARAMIQOQSNNKNCPOKEQVSSSYLGS 570
Qy 264 DFEKVTGR-IDKNVSPGARHPVLAAYPIVHVMENILSKNEDOSTQDSETRTISKNT 322
Db 571 NGSNLSGRSLDES---EQIPLKD-----FTENNNNLKTGDGLSSSVETI 613
Qy 323 STSRHTSEVGNMAEVH-ANTSTRTHTSEVGNMAEVHVAIDSLSLAGERTWAETWGL 381
Db 614 EKEVSEKKLDGSTEKELVPLSTDTTINSSLGNEDSIYSLDADASEMLTIVPLMEI 673
Qy 382 NTA-----DPAFLNANIRVNTGAPLYNVLPPTSLVGNQTLAT 422
Db 674 KTTPEKYEVVISEVSYSTSYEDNTVAMPPOVEY---TSPPMND-PENSL---ND--D 721
Qy 423 IKAKENOLSOILAPNNVPSKNLAPIALNADDFSPPTITM-----NTNOFLELEKT 474
Db 722 YEKHDLKSTLAA-----LAPATFKDAEVEVAGVTSCLTSTSGHTNITHTSKET 773
Qy 475 KQLRLDQVYGNATATYVNFENGRRVVDGNSMSEVLPOIQTETARIIFNGKDLNVERRI 534
Db 774 KOVS-DLDEESTENTYFENENTG---DENKQSKNFPGVANSTDKSTEDNDI---EKYF 824
Qy 535 AAVPSPDLETTKPDMLTKEALTAFGFENPENGNYQYQKITERDFNF--DOQTSQN--- 590
Db 825 SAINTTN---VTGDSCEDIETASVNEE---NLRYCEKMNEMKSSGDECVAKONDG 877
Qy 591 IKNOLA-----ELMATNITYVLDKIKLNKMNILIRKRFHRRNIAVGADESUYKE 643
Db 878 SKTQISFSTDPDNFQESNDMTFFSSTK-----YKVRNSDLEDESILKE 922
Qy 644 AHR-EVIN-----STEGELLNIDKDIRKILSGYIEIEDTEGLK 682
Db 923 LTKAEVVDKLDDESEDSYEDDYADPEPGNDEGSNENIVKGTKK-----DTLGIV 972
Qy 683 EVIDRDMMLISSLRQGGKTFIDFKKNDKPLXISNPNKVVNVYATKENTIIIN 738
Db 973 EPENEKVN-----KVHEEETLEFANVSSVNVQNMKMDHTDVIN 1010
RESULT 5
SMC3_YEAST STANDARD: PRT: 1230 AA.
AC P47037;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CHROMOSOME SEGREGATION PROTEIN SMC3 (DA-BOX PROTEIN SMC3).
GN SMC3 OR YJL074C OR J1049.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-W303;
RX MEDLINE-97474309; PubMed-9335333;
RA Michaelis C., Cloos R., Nasmyth K.;
RT "Chesins: chromosomal proteins that prevent premature separation of
RT sister chromatids";
RL Cell 91:35-45(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Rose M., Koelter P., Entian K.D.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [3]

RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RA Sor F.J.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC PART OF A CHROMOSOME CONDENSATION MOTOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y14278; CAA74655.1; -;
DR EMBL: Z49349; CAA89366.1; -;
DR EMBL: X88851; CAA61313.1; -;
DR SGD: S0003610; SMC3.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02483; SMC_C; 1.
DR Pfam: PF02463; SMC_N; 1.
KM Mitosis; ATP-binding; Coiled coil; Nuclear protein.
FT NP_BIND 32 39
FT DOMAIN 172 482
FT COILED COIL (POTENTIAL).
FT COILED COIL (POTENTIAL).
FT DOMAIN 685 1041
FT COILED COIL (POTENTIAL).
SQ SEQUENCE 1230 AA; 141336 MW; B152D88F7780341F CRC64;
Query Match 4.48; Score 173.5; DB 1; Length 1230;
Best Local Similarity 20.04; Pred. No. 0.24; Indels 247; Gaps 39;
Matches 186; Conservative 137; Mismatches 358;
Qy 22 GNLEVTQAEVQENRLNSESSESSGGLGYFSDLPFQAPVAVT--SSTTGDLSPSEL 79
Db 201 GELNSKLSMEQERKELEYNLEENRKIIYFTLYDRLEVINOMERLDDYN----- 254
Qy 80 ENISENOYFO-----SAIWSGFIYKKSDEYTPASADNHVTMMVDDO 123
Db 255 NTVYSSQYIOLDKREDMDIDOVSKKLSI-EASLIKINATDLOAKLRSEISOKLTNV 313
Qy 124 EVINK-----ASNSNKRLEKGRLOYIK-IQYQRENTFEGFLPKLYTDSQKKVEIS 176
Db 314 NVKIKDVQOQIESNEEQNLDSATLKEIKSIIEOKKOLKILPRYOELT---KEEAMV 369
Qy 177 SDNLQPELKOKSSNRKRRKSTAGPTVPDRDNDGIPDSLEVEGYTVQVKKRTFLSPWI 236
Db 370 --KQLASLQOKNDLILKKEVYARFKSKDERDWMHSEIE-----ELKSS-----I 414
Qy 237 SNIEH-KKGLTKYSSPEKWTASDPYSDEPKYGRIDKNVSPGARHPVLAAYPIVHVM 295
Db 415 QNLNELESQLOMDRTSLKQYSAID-----EIEELIDLSINGPPTKGL-----EDFDS 463
Qy 296 ENILSKNEDOSTQNTSETRTISKNTSRTHTSEVGNMAEVHVAIDSLSLAGERTWAETWGL 355
Db 464 ELIHLKOKLSLDTREKELMRKEQLQTVLETLSDVQON-QRVNNETMSH-----LAN 517
Qy 356 AEVHVAIDSLSLAGE-----RTWAEFTMG-----LNTADTARLNA 391
Db 518 GIINWKEITELKLSPEVEFTIGELIKVNDKXKTCAYIGNSLFLHVDTEETATILM 577
Qy 392 NIRVYNTGAPLYNVLPPTSLVGNQTLATIKAKENOLSOILAPNNVPSKNLAPIALN 451
Db 578 NELYRMKKGAVTF--IPLNRLSDSDVFPSTNTTQIOFTPLIKIKKEPFEKA----- 630
Qy 452 AODDESSPTIMNNQFLELEKTKQLR--LDTPQV-IGNLAT--YNEENGRVAVDTGSN 505

```

Db 631 VKHVFQKTVVKGDLGGGLAKKHKLNATITLDDGRADKRGVLITGGLDQHKRTRESLKN 690
Oy 506 MSEVLQIOETARTIENGKDLVVERIAAVPS-----DPLETT 546
Cc 506 MSEVLQIOETARTIENGKDLVVERIAAVPS-----DPLETT 546
Db 691 LNSRSQHKKILEEDLVNRELNDITKIDVGNINRKYNSNRESVLTINIEVYRTSLNTK 750
Oy 547 KPD-MTLKALK-IAGFNEPNCNLQYOGKDIPEFN-----PDQOSONIKNOLA----- 596
Cc 547 KPD-MTLKALK-IAGFNEPNCNLQYOGKDIPEFN-----PDQOSONIKNOLA----- 596
Db 751 KNEKLLESLSNLIILKEKLNTRTFADQEKLMFTFENDLLQGEDSELSKEKRELESRTK 810
Oy 597 -----ELNATN-----IYVLDKIKLAKKNILIRDKRFHYDRNNIANG----- 635
Cc 597 -----ELNATN-----IYVLDKIKLAKKNILIRDKRFHYDRNNIANG----- 635
Db 811 EISAANKINTISDALEGITTTID--SLNAELESKLIPQENDLESKMGDAFIFGLD 868
Oy 636 -----ADESVYKEAHRVINSST-----EGLL--N 659
Cc 636 -----ADESVYKEAHRVINSST-----EGLL--N 659
Db 869 ELKELOLEKESYEHQENAVLEIGTVOREIESLIAETNNKLLERANNQOIRLLKLDN 928
Oy 660 IDKDIRKISGYVEIEDTEGLKE-----VINDRYDM-----LN-----I 694
Cc 660 IDKDIRKISGYVEIEDTEGLKE-----VINDRYDM-----LN-----I 694
Db 929 FOKSVKTKMKTTLVTRRELQORIREIGLPEDALVNFSDITSQOLLRLDMNTEI 988
Oy 695 SSLRQDGK-TFIDFKKYNDK-----LPLYSNPKYKVVAYATKENT 735
Cc 695 SSLRQDGK-TFIDFKKYNDK-----LPLYSNPKYKVVAYATKENT 735
Db 989 SGLKNVNRKRAFENFKFNERKDLARASELDESKDSIODLIYKQKQVNAV----DST 1044
Oy 736 IIMPSENDT-----STNGIKKILIFSK 758
Cc 736 IIMPSENDT-----STNGIKKILIFSK 758
Db 1045 FOKVSENFVAFERLVRPGTAKLIIHRK 1072
RESULT 6
Y338_MYCGE
ID Y338_MYCGE STANDARD: PRT: 1271 AA.
AC P47580;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOHETICAL LIPOPROTEIN MG338 PRECURSOR.
GN MG338.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sultoon G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 1023-1114 FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bort K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
Cc -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
Cc ANCHOR (POTENTIAL).
Cc -1- SIMILARITY: BELONGS TO THE MG307 / MG309 / MG338 FAMILY.
Cc -----
Cc This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc between the Swiss Institute of Bioinformatics and the EMBL Outstation -
Cc the European Bioinformatics Institute. There are no restrictions on its
Cc use by non-profit institutions as long as its content is in no way
Cc modified and this statement is not removed. Usage by and for commercial

```

```

Cc entries requires a license agreement (See http://www.isb.ch/announce/
Cc or send an email to license@isb.ch).
Cc -----
DR EMBL: U39715; AAC71563.1; -.
DR EMBL: U01809; AAD12341.1; -.
DR TIGR: MG338; -.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Lipoprotein; Membrane; Signal;
KW Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 1271
FT LIPID 27 27
FT SPOURCE 1271 AA; 142492 MW; FCE042067310A70 CRC64;
SO SPOURCE

Query Match 4.4%; Score 173.5; DB 1; Length 1271;
Best Local Similarity 18.4%; Pred. Nd. 0.25;
Matches 181; Conservative 143; Mismatches 337; Indels 321; Gaps 42;

Oy 14 STIVSSTGNLEVIQ-----AEYKQENRLNE-----SESSQGLGYFFSDNLFQAP 61
Cc 14 STIVSSTGNLEVIQ-----AEYKQENRLNE-----SESSQGLGYFFSDNLFQAP 61
Db 317 SKVLISKNNILSVLKYVNLGSAAYIDQYHYLLNNTKELTTTITTTTGGTSSNLPDK 376
Oy 62 MVTYSS-TTGDLSIPSELENIPSENOYFQSAIWGFIKYKSDDEYFATSDNHTMYV 120
Cc 62 MVTYSS-TTGDLSIPSELENIPSENOYFQSAIWGFIKYKSDDEYFATSDNHTMYV 120
Db 377 FIKSSSATVYMKMSAMTKSQDEVTSDDNGFN--VKSEFLKINPLSSSGSDMSNTQSEFWK 434
Oy 121 DDQEVINKASNS--NKIRLEKG-----RLYQIKIYQRENPTEKGLDKLY 164
Cc 121 DDQEVINKASNS--NKIRLEKG-----RLYQIKIYQRENPTEKGLDKLY 164
Db 435 QVQALNNSQATATTFDAVRMSNSSQAQVYVSNLVSLSKTKQKQKQKPYVYRGDAIY 494
Oy 165 -----WDSQKKKEVISSDNQLPELKQKSSNRKRSATGAPVPPDNDNGIP 213
Cc 165 -----WDSQKKKEVISSDNQLPELKQKSSNRKRSATGAPVPPDNDNGIP 213
Db 495 AFHIDGNYLENSPKNRKFQAEVLLMRFLQGQTNFSK-----DNVS 540
Oy 214 DSLEVEGYTVDK--NKRTFLSPWI-----SNIH 240
Cc 214 DSLEVEGYTVDK--NKRTFLSPWI-----SNIH 240
Db 541 FSVYDLFQNSSEFRMANRNTTKLYTALTMLENGTSNNQKQYCDLAKKLNNTNLS 600
Oy 241 E-----KGLTKYKSSPKKMSASDPYSDPEKVTGRIDK-----NVSPE 279
Cc 241 E-----KGLTKYKSSPKKMSASDPYSDPEKVTGRIDK-----NVSPE 279
Db 601 ETIKQODFNNSLSQIKSSSESYIKAAKNLNFPTDLANIEQLQIVDRANNYIKLQKE 660
Oy 280 ARHPLV-----AAYPIVHDMENILSKREDQSTQNT-----DSE 314
Cc 280 ARHPLV-----AAYPIVHDMENILSKREDQSTQNT-----DSE 314
Db 661 AKESSTIGQPLPYKRRANDSYSLAKFPNN-----MSDSSAOTLLTKTTAAITSDNE 715
Oy 315 TRITSKNTSTSRHTSEVHGNAEYHANTSTSRHTSEVHG-----NAEYHAVAIDHSLS 368
Cc 315 TRITSKNTSTSRHTSEVHGNAEYHANTSTSRHTSEVHG-----NAEYHAVAIDHSLS 368
Db 716 P-TREKNNQTLKLTLEVENKAKAKELVEKKATYVSSSOYSLIRLKSSQDLNDLDDILS 774
Oy 369 LAGERIWAETWGLNTADTARLANIRYVNT-----GTAPIYVNLPTTSLV 413
Cc 369 LAGERIWAETWGLNTADTARLANIRYVNT-----GTAPIYVNLPTTSLV 413
Db 775 ILTFD-----SGIRTGIVANIFKMWYFKNTSPFNNDTSNKEKLGCFSPFNLDVKKALY 828
Oy 414 LGRNQTALATKAKENQSLIAPNNYPP--SKNLALIALNAQDDFSPTITMNYQFL- 469
Cc 414 LGRNQTALATKAKENQSLIAPNNYPP--SKNLALIALNAQDDFSPTITMNYQFL- 469
Db 829 IRSMQNLN--SKER-----FGYKDKGVSNTSTMTQMOKKAAQSHSTSSVNOJTL 877
Oy 470 -----ELEKTKQRLPDTQYVGNL--ATNFENGRRVDTGSMSEVLQIOETT 517
Cc 470 -----ELEKTKQRLPDTQYVGNL--ATNFENGRRVDTGSMSEVLQIOETT 517
Db 878 DLAKKAEKELEDNQ--DAEYKYMRFLQALMLVANG-----AONYKNLQOALPIG 928
Oy 518 ARIIFN--GKDLNVERRIAANVPSDPLETTKEDMTLKEALKIAGF-----FNEBNGN 568
Cc 518 ARIIFN--GKDLNVERRIAANVPSDPLETTKEDMTLKEALKIAGF-----FNEBNGN 568
Db 929 TRAFVSTVGYDKN-----PSATVYQKTKSTSSANENPNRFLQN 969
Oy 569 LQY-QGQDITEFPNPOQTSQNIKNQALNATNITYVLDK-----IKNAKNILIRD 622
Cc 569 LQY-QGQDITEFPNPOQTSQNIKNQALNATNITYVLDK-----IKNAKNILIRD 622
Db 970 PNTQGESEIMFN--DKQPTIOPDSL--LESNTYRFTDEPNNSVALSNKSO-GSSD 1032
Oy 623 KRPHYDRNNIANGADESVKEAHRVINSSTEGLLINDIKRIL-----SGYVEIED 677
Cc 623 KRPHYDRNNIANGADESVKEAHRVINSSTEGLLINDIKRIL-----SGYVEIED 677

```



```

FT CONFLICT 1568 1568 D -> V (IN REF. 1).
FT CONFLICT 1630 1646 DLKOLDHYTKVEMLN -> SEARSLYKGGNVD (IN
FT CONFLICT 1698 1704 REF. 1).
FT CONFLICT 1725 1737 MISSING (IN REF. 1).
FT CONFLICT 1754 1757 TLOLOMOUNSRNG -> NTJANGTKREW (IN REF.
FT CONFLICT 1777 1777 1).
FT CONFLICT 1788 1788 FDDE -> LMM (IN REF. 1).
FT CONFLICT 1825 1825 D -> E (IN REF. 1).
FT CONFLICT 1882 1882 R -> T (IN REF. 1).
FT CONFLICT 1902 1904 S -> D (IN REF. 1).
FT CONFLICT 1928 1904 S -> W (IN REF. 1).
SQ SEQUENCE 1928 AA: 223634 MM: 6F54C7611F43DC9F CRC64:
FWK -> NSGKRDLADL (IN REF. 1).

Query Match 4.38; Score 166.5; DB 1; Length 1928;
Best Local Similarity 19.28; Pred. No. 0.97;
Matches 162; Conservative 145; Mismatches 291; Indels 247; Gaps 39;

OY 20 STGNLEVOAEVROENRLNESSESSOGILGYFSDNLFQAPMVVTSSTGDL 73
DB 894 TVNELETTQDLNOEKENLNKKNESLNKRVTSSTLOKOFDDL-----VSEKDEIS 944
OY 74 IPSEL-ENIPSENOYFQSAIWGFIKKKSDEYTFAT--SADNHYVMWDDQEV-INKA 129
DB 945 REKLEVAQNEEAHQKIQ-----GLQETIREREATLEKLSKNNELKQISDLNCDISKE 999
OY 130 SNS-----NKIRLE-----KGRLOYIKIOYORENPTKEGLDFKLYWTD----- 167
DB 1000 OSSOSLIKESKLEKLENIKRLKDVINSKEEIKSFNDKLSSESDIDIKITVLEKKNCIA 1059
OY 168 -----SON-----KKEVLSSDN-----LQELKOKSSNRKRSSTAGPT 203
DB 1060 MSRLQSLVTNSDLRSNENFKEKALNMQKNKESLKMKEKIDNHKKELATFS--- 1116
OY 204 VPDNDNGIDPSLEVES-YTVYKNNKRTFLSPWISNITHEKGLKYKSPEKSTASDPY 262
DB 1117 -KQRD-----DAVSEHKITAELEKTRI-----QLTEKYSNOK--IKREY 1154
OY 263 SDEKTYGRDKNVSPARHPVAAYIVIVDMENILSKNEDOSTONTSETTISKNT 322
DB 1155 SNFORE-----KEOEKRRNSLV-----ESLNDSKIRKELEAR-----LSQEI 1192
OY 323 STGRTHSEVHGAEVANSTSRTHS-----EVHGNAEVH 359
DB 1193 SLNQYLNRKRSNGS-VETNISSTRSTSYSDDPDKEDIIKKYYDLOLAFETRINENE 1251
OY 360 AVAIDHSLAGERTMAET-MGINTADTARLMANIRYVNTGTADIVNLPPTSLVLGKNQ 418
DB 1252 ---TEERKNLISRLRFETRLASSFEDQKIKQMKKIKKLIDMDPSIPLDSIL---NE 1305
OY 419 TLATIRAKENQSOILAPNNYIPSKNLAPALNAQDFSSPTITMANYNOFLEKTKQLR 478
DB 1306 PLNCPKEDDKIMLEVYL--KROLDIETRAHYD-AENALISALSKRKRIQ--GESS 1360
OY 479 LDIDOVYGNATJNFENGVRVDTGSMWSEVLPQIET---TARIINGKDLNIVERRIA 535
DB 1361 LSSSDI-----KLFESSEERKVSLEDKLTMPLRDRPTNLPVGDIIKRRDISKYEIEIR 1416
OY 536 AVNPSPLETTKPDMLTKALKIAFGNEPNCNLOY----- 571
DB 1417 YRK-----LENYK-----NESNGKLSQLTLDLRQSKSEKALLSEOLDRLQK 1462
OY 572 -----QKQITEPDEFPDQOTSONIKNL-----AELNATNIYTVLDK 609
DB 1463 DLESTERQKELSLSTIKQKQOFCNCMDLQGNELRLREHIALKQAEEDKNNASTIEK 1522
OY 610 IKL-NAKMNILIRDKRFHYORNNIANGADESVYKAEHREVINSSTEGLLNIDIDIKIL 668
DB 1523 LKQONKQKELIMEREM--ERNDSMDLOLETLE-----LKRQDVAKKIL 1565
OY 669 SGYIVE-----IEDTEGLKEVINDRYDMLNISSLRODCKTFIDFKKYNKDLPLYTSPN 721

```

```

DB 1566 SDDLHLKRLSAVEDRSQYTDIEINRLKEELN-CSLKAETNLRKEFATLKYKLETTNDS 1624
OY 722 NYKNV 726
DB 1625 EAKIS 1629

RESULT 8
MYS2_DICTI STANDARD; PRT: 2116 AA.
AC P08799;
ID MYS2_DICTI
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN II HEAVY CHAIN, NON MUSCLE.
GN MHCA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelidae; Dictyostelium.
OX NCBI_Taxid=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87092266; Pubmed=3540939;
RA Marlick H.M., de Lozanne A., Leinwand L.A., Spudich J.A.;
RT "Conserved protein domains in a myosin heavy chain gene from
RT Dictyostelium discoideum.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RN [2]
RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
RX STRAIN-AX2;
RX MEDLINE=90353583; Pubmed=2387408;
RA Lueck-Vielmeier D., Schleicher M., Grabatin B., Wippler J.,
RA Gerisch G.;
RT "Replacement of threonine residues by serine and alanine in a
RT phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
RL FEBS Lett. 269:239-243(1990).
RN [3]
RP PHOSPHORYLATION SITES.
RX MEDLINE=88112226; Pubmed=2828113;
RA Magle G., Noegel A., Scheel J., Gerisch G.;
RT "Phosphorylation of threonine residues on cloned fragments of the
RT Dictyostelium myosin heavy chain.";
RL FEBS Lett. 227:71-75(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX MEDLINE=95345066; Pubmed=7619795;
RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutcliffe H.M.,
RA Rayment I.;
RT "X-ray structures of the myosin motor domain of Dictyostelium
RT discoidium complexed with MgADP, BeFx and MgADP.ALFA-";
RL Biochemistry 34:8960-8972(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX MEDLINE=95345067; Pubmed=7619796;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
RT truncated head of Dictyostelium discoidium myosin to 2.7-A
RT resolution.";
RL Biochemistry 34:8973-8981(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=96206189; Pubmed=8611530;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II).ADP.vanadate complex of the
RT Dictyostelium discoidium motor domain to 1.9-A resolution.";
RL Biochemistry 35:5404-5417(1996).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX MEDLINE=97452580; Pubmed=9305951;
RA Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT "X-ray structures of the MgADP, MgATPgammAS, and MgAMPPNP complexes
RT of the Dictyostelium discoidium myosin motor domain.";
RL Biochemistry 36:11619-11628(1997).
RN [8]

```

RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
 RA MEDLINE:98070605; PUBMED:9405148;
 RX Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.,
 RT "X-ray crystal structure and solution fluorescence characterization
 of Mg₂'(3')-O-(N-methylanthraniloyl) nucleotides bound to the
 RU dicyostellium discoidium myosin motor domain.";
 RL J. Mol. Biol. 274:394-407(1997).
 CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
 CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
 CC -1- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
 CC INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
 CC LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
 CC (MLC-2).
 CC -1- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
 CC CORTEX.
 CC -1- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEMOROSIN (LMM) AND 1 HEAVY MEMOROSIN (HMM). IT CAN BE FURTHER
 CC SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -1- DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
 CC THE ACTIN-ACTIVATED ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: DICTYOSTELLUM MYOSIN II HAS NO K(2)EDPA ATPASE
 CC ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CTS AT THE SH-1
 CC POSITION (688).
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M14628; AAA3227.1; -
 DR PIR: A26655; A26655.
 DR PIR: S00250; S00250.
 DR PDB: 1MMA: 03-DEC-97.
 DR PDB: 1AMD: 17-AUG-96.
 DR PDB: 1IMG: 03-DEC-97.
 DR PDB: 1MMN: 03-DEC-97.
 DR PDB: 1MND: 17-AUG-96.
 DR PDB: 1MNE: 17-AUG-96.
 DR PDB: 1VOM: 23-DEC-96.
 DR PDB: 1LVK: 28-JAN-98.
 DR DictyDB: DD01008; mhca.
 DR InterPro: IPR000048; IQ.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00612; IQ; 1.
 DR Pfam: PF00063; myosin_head; 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR ProDom: PD000355; myosin_head; 1.
 DR SMART: SM00015; IQ; 1.
 DR SMART: SM00242; MYSC; 1.
 DR ProSite: PS00096; IQ; 1.
 KW Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
 KW Calmodulin-binding; Methylation; Alkylation; Phosphorylation.
 FT DOMAIN 1 761
 FT DOMAIN 762 791
 FT NP_BIND 817 2116
 FT NP_BIND 179 186
 FT DOMAIN 638 752
 FT DOMAIN 738 752
 FT MOD_RES 130 130
 FT MOD_RES 678 678
 FT MOD_RES 1823 1823
 FT MOD_RES 1833 1833
 FT MOD_RES 2029 2029
 SQ SEQUENCE 2116 AA; 243871 MW; 2FC370B1E56A1 CRC64;

Query Match 4.1%; Score 162; DB 1; Length 2116;
 Best Local Similarity 19.2%; Pred. No. 1.8;
 Matches 156; Conservative 127; Mismatches 296; Indels 232; Gaps 36;
 2 KRRKVLPLMALSTILVSTGSLVLIQAEVQKQENRLNSESSESGGLGYFSDINFGAP 61
 1200 QKKKVELDLEDSQAQAEETAKQALDKLKKLEQELSEVQTO-----LSEANNK-- 1249
 62 MYVTSSTGDSIPSELEINISENOYQSAWSGFYKKKDETFATSDNHTMWD 121
 1250 NVNSDSTNKL--ETSFNNKLLEAEQK-----KQALEKKRLGSELEKHNQEOLE 1300
 122 DEEVINKASNSKIRLEKQRLQIKYQORENPTKGLDFKLYWDSQNKKE----- 173
 1301 EKKKQESNEKKKVLDEK-EVELNDQIEEVAASKAV-----TEAKKSELEDEIKR 1353
 174 ---VISSDNLQPLPELKOKSSNSRRKSTSAQTPVDRDNDGIPDSLEVEGYTVDKNR 229
 1354 QYADVSSRDKSVQDLTKLQAKNEELRNPA-----EEAQGLDRAERS-----KKKA 1400
 230 TFLSPWISNIEKKGLKRYKSPKWSADPSDFEYVGRID--KNVSPEARHPLVA 286
 1401 EFDLEAVAKNLEE--TAKKAKAKAKKAE--TDYNSTSELDKAKNVSEQ----- 1449
 287 APTVHVDMENILSKNEDOS-----TONTSEPTISKNTSTRTSEVGNAEVHAN 341
 1450 ---YVQIKRL-----NEELSELRSVLEADERNSAIKAKKTAALESLELDEIDANN 1500
 342 TSTSTRTSEVGNAEVHANVALDHSLSLAGEETMAETAGLNTADTARLANRINYNGTA 401
 1501 AKAKAKERSK--ELEVRVALEESLE-----DKSG-----TVNVEFIRKDA 1540
 402 PIYVNLPTSLVGLKQNTLATIKAKENQSLIAPNNYPSKLPALMAODDEFSPI 461
 1541 ELDIDLRARD-----RETFSRKSDSK-----KN----- 1565
 462 TNNVQFLEKTKQRLDFTDQVYNIAATYVNEGSRVAVDGSNMSEVLPOIO-ETTARI 520
 1566 --TRKQFADL-----AKVEAQREVVITD---RLKKLESIDILDTQDTETKSR 1613
 521 -TFNCK--DNLVERRRIAANVPSDPLETTKPDMLKELKIAFGFNPNGNLOQSGDI 576
 1614 KTEKSKKLEQTLERRAAEBGSSRAAD-----EELRKQVW-----QEV 1652
 577 TEFDFNFDOQ-----TSQNIKQALAEINATNTYTVLDKIKLNAMNILLRDKRPHYDRN 630
 1653 DELRAQLDSERAAALASEKKIKSLVAE-----VDEKQLEDELLARDKLVAKRA 1703
 631 ---NIAVGADESVAHREYVINSSTEGILLINDKDIRKILSGYIETE----- 676
 1704 LEVELEEVNDQLEEEEDSELEDSKRRLLTTEVE-DIKK---KYDAEVEQNTKLDEAKKK 1759
 677 ---DREGLEKVINDRYDMINISS-----LRQDGKFFI-----DFKKYDK 713
 1760 LTTDDVDTLKKQLEDEKKNLSEERAKKRLESENEDFLAKDAEVKNSRAEKDRKRYEKD 1819
 714 LPLYSINPNYKVNVAATKENTINPSENGD 744
 1820 L-----KDKRYKLNDAARIKTQTEIGAANKLED 1846
 RESULT 9
 RBPL_PLAVB STANDARD; PRT: 2869 AA.
 RBPL_PLAVB 000798:
 01-APR-1993 (Rel. 25, Created)
 01-APR-1993 (Rel. 25, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
 RBPL.
 Plasmodium vivax (strain Belen).

CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_Taxid=11273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites";
RL Cell 69:1213-1226(1992).
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -1- SUBUNIT: HOMODIMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M68097; AAA29743.1; -.
DR HSSP: P36956; IAM9.
KW Malaria; Receptor; Signal; Transmembrane.
FT SIGNAL 1 17
FT CHAIN 18 2869
FT DOMAIN 18 2807
FT TRANSMEM 2808 2826
FT DOMAIN 2827 2869
FT SITE 1030 1032
FT SITE 2599 2601
SQ SEQUENCE 2869 AA; 330213 MW; B9DBA442205BECF CRC64;

Query Match 4.1%; Score 162; DB 1; Length 2869;
Best Local Similarity 18.1%; Pred. No. 2.8;
Matches 150; Conservative 150; Mismatches 319; Indels 210; Gaps 38;

OY 5 KVLPLMALSTILVSTGNEVIOAEVKOE-----NRLNESESSSGGLGYTSD--- 55
DB 1909 KTLINDISECSVL---NNVSVYKWKESKHADYRDRDANSWESMVTLANFLYSDAKI 1964
OY 56 ---LNFQAPWVYVSTTGDLSPSELENIPESENOYFQSAIWGFIYVKKSDVTFATSA 112
DB 1965 SSGMEFNAEKSNFKTLELEITSV-----ISNSNEL-----LKKEQ----- 2002
OY 113 DNHVTMVDQEVINKASNSKIRLEKGRLYQIKIYQRENPTEKGLDFKLYWTDSONKK 172
DB 2003 -----DSNDVIQKRESEQLAKDATDIYV-IRLKNE-----FNEKLEAKNKE 2045
OY 173 EVISSDNLQLELKKOKSSNRKKRSTAGPTVPDRDNDGIPDSLEVGYYVAVKNNKTFEL 232
DB 2046 EYVS-----EKVREALKRLS-----QVGGIRCHFPNFRLL 2076
OY 233 SPWISNIHE---KKGITFKKSSPEKWTASDPYSDFEYVGTGRIDKNVSPARHPLAAY 288
DB 2077 ---DNTLEENLKKMTYTR--DKKSERBSGLQEMENEMNTYSNIT----- 2118
OY 289 PIVHVDMENTIL---SKNDQSTONTSETRTISKNTSTSRTHSEHGNAEVHANTST 344
DB 2119 ---QLEGIVSAGEKEDEIKERSENEEMRNISEKISTIDSKVIMENSTID----- 2166
OY 345 SRHTHSVHGNAEYHAVAIDHSUSLAGERTVAETGTLNADTARLANIRYVNTGTADPIY 404
DB 2167 ---ELYKLGKNCQAHMISL---ISYTANMKTSKLIMINKENTKCVADIKDNSSSTD 2220
OY 405 NVLPSTSLVGNKOTLATITAKENQSLIAPNNVYPSKNLA---PLALNADQDFSSPTPI 461
DB 2221 GYVETLKGFGSKLTSS-----ASEIVONADPTV-SVNFPAKHEKESLNAIRIKKELY 2272
OY 462 TNNYNQLEL-----EKTQRLRDTQDVGNINATYFNENGVRVADTGSNNSEV 509

DB 2273 LFHONSDSIYEGGVONMALYDLKLINEKREMDLEYRNISSETKLQKHSHSTDFVKPMIEL 2332
OY 510 LPQIOETARIIFNGKDLNLVERIAVNPSPLETTKPDMTLEAKI-----AFGRNEP 565
DB 2333 HKGNMET-----NKKSLLEKREKRLKSVN--DHMSMEAEV-IRNGKLYPESVONINNI 2383
OY 566 NGNLOYOGKDTTEEDPNE--DOOTSQNKLOAEL-NATNITYVLVDIKLAKANNILIRD 622
DB 2384 YSVTEAEVKTLTEEDROGDVNYQIVEEHKKQFSLIDRTN--ALMDIEIRKEN----- 2436
OY 623 KRPHY---DRNNIAGADESVYKKAHREVINSST--EGLLNIT--DKDIRKILSGYIVE 674
DB 2437 ---NYNLMEVNTETIHRVNDIEIKTKNLYQAKTEVEQIILEENKONDMQNIPLKKVSI 2493
OY 675 IEDIEGLKE---VINDRYDM-----LNTSLRQDKTITDPRKKNDRPLYSN--- 720
DB 2494 IEYEVNKKKKESLINDLYEGERLLKIGEHIDELIKRNVETLSYEIDQKEMMSKNILE 2553
OY 721 ---PNYKVVAVYVTKENTTIIPS---ENGDTSTNGIKILIFSKKG 760
DB 2554 KSKMMMY-TSYELERANFINDAKOIKDDTIINSVLEAL-QKRG 2600

RESULT 10
YD86_SCHPO
ID YD86_SCHPO STANDARD; PRT; 1957 AA.
AC Q10411;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHECAL 222.8 KDA PROTEIN C1F3.06C IN CHROMOSOME 1.
GN SPAC1F3.06C
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_Taxid=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z70690; CA94624.1; -.
DR Hypothetical protein.
KW
SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

Query Match 4.1%; Score 161; DB 1; Length 1957;
Best Local Similarity 17.4%; Pred. No. 1.8;
Matches 162; Conservative 150; Mismatches 319; Indels 298; Gaps 37;

OY 14 STIIVSSSTGNEVIOAEVKOEENRLNESESSSGGL-----GYFSDLN----- 57
DB 482 SALIDKDDQELNLRBOQIKKQKVSSTQSSLOSLORILNEKKKHEYESQLNDELKGL 541
OY 58 -----FQAPWVYVSTTGDLSPSELENIIPS-----ENO 87
DB 542 QTEISNSEHSSQSLTAAEKAAVATNNELSEKNSIQTLCKNFQEKLAQSVQMLKENE 601
OY 88 YFQSAIMSGFIYVKKSDPTATSDNHYVTVMDQEVINKASNSKIRLEKGRLYQIKI 147
DB 602 QNFSSLDTSFKKLNESSH-----ELENHQTIT--TKQDKRHSKLOQDLERANFEOKES 654
OY 148 QYOREN-----PTKGLDFKLYWTDSONK-----KEVISSDNLQLEP 184


```

Db 655 TLSDNNDRKRLKLEESNNSLIKQEDVDSLEKNIOQLKEDLKKSEALFSTLEAKN 714
Qy 185 LKOKSSNRKKRSTGAGTVPDRNDGIDPSLEVEGYTVYVKNKRTFSPMISNHEKKG 244
Db 715 LREVIDNLKGRHETL-----EAQRNDLHSSLS-----DAKNTAIISSLELT----- 755
Qy 245 LTKYKSSPEKSTASDPYSDPEKVTGRT-----DKNVSPEARHPLVAAPYV----- 291
Db 756 -----KSS-----EDVKRLTANVETTLQDASKAMQOSTPISVNSYOSTISNLVHELR 800
Qy 292 --HYVMEIIISKNEDOSTQONDSETRISKTSPSRHTSYVHNAEYHANTSIRHT 349
Db 801 DDHVMQSONNTLSESESKLTDCENL-----TQONMTLIDVQKLMKHVQESKVSSEL 855
Qy 350 SEVHNAEYHAVAIDHSLSLA-----GERTMAETMGLN-----TADTALNANIRVYN--- 397
Db 856 KEVKNKSLDLKLNLSLVAISDNDQILQLAELSKNVDLSQESADLNSGLKLEAK 915
Qy 398 -----TGAPYIVLPTTSLVGLKNOFLATIKAKENOLSOILAPNNY 440
Db 916 QLLHTENEELHRLDKLGLKLTIEE---SKSSDLGKLT---AROEISMLKEEN--- 964
Qy 441 PSKNLAPIALNADDESSPTITMANNQFLELEKTKQLRDLTDQVYGNATYNE----- 494
Db 965 MSQSAITSVSKLDET-----LSKSKLEADIEHLNKNVSEVEVERNALLA 1011
Qy 495 -NGRVAVD--TGSNMSEVLPOIOTTAIIIFNGKDLVERRIAVNPSPLETTKPD 550
Db 1012 SNERLMDLKNNGENIASLQETIEKKRAE-----NDLQ--SKLSV 1050
Qy 551 TLKELAKTAFGPNENGMLQOGKDIETEDNR---PQOQTSQNIKNOLAELNATNIYTV 606
Db 1051 VSSEYENLLISSQTNKSTLEKTNQKYLEKVNQKLDEKQDRN--ELEEL--TSKYGK 1106
Qy 607 LDKIKLAKMNLIL--FDRKRF-----YDRNNIIVGADESYYK 642
Db 1107 LGEENAOJKDELALRKSKKQHDICANFVDDLKKSADLEOLINEKVELIVSLEOS--- 1163
Qy 643 EAHREVINSSTEGT-----LLNIDKIRKILSGYIVEIEDTEGLEKEYI-----NDRY 689
Db 1164 -----NSNDEALVEERSDLANRLSDMKKSL-----DSDNIVSIRSDLVAVNDEL 1209
Qy 690 DML-----NISLQODGKTIFD-----FKYNDKILPLTISPNKRVNYAVATK 732
Db 1210 DTLKKDKSLSTQYSEVQCDRDLDLSLKGESESPKNYAVASLRELCSTSEIDVPSIJD 1269
Qy 733 ENTIIINPSENGDTSPNGIKKILIFSKKGY 761
Db 1270 DNFVFNAGNFSELS-----RLTVLSLENY 1293

```

```

RESULT 11
USOL_YEAST STANDARD: PRT: 1790 AA.
AC P25386:
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTRACELLULAR PROTEIN TRANSPORT PROTEIN USOL.
GN USOL OR INT1 OR YDLO58W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_Taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RA MEJLIME-91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol, is required for intracellular
protein transport in Saccharomyces cerevisiae."

```

```

RL J. Cell Biol. 113:245-260(1991).
RN [2]
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -!- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: X54378; CAA38253.1; -.
DR EMBL: L03188; AAB00143.1; -.
DR EMBL: U53668; AAB6659.1; -.
DR PIR: A38455; A38455.
DR HSSP: P80220; IDIP.
DR SCD: S0002216; USOL.
DR InterPro: IPR002017; Spectrin.
KW Transport; Protein transport; Golgi stack; Cytoskeleton; coiled coil.
FT DOMAIN 1 724
FT DOMAIN 725 1790
FT FT 465 1790 COILED COIL (POTENTIAL).
FT DOMAIN 991 1790 CHARGED (HYPER-HYDROPHILIC).
FT FT 1172 1786 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 847 847 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 924 924 E -> K (IN REF. 2).
FT CONFLICT 1253 1253 V -> I (IN REF. 2).
FT CONFLICT 1319 1319 I -> V (IN REF. 2).
FT CONFLICT 1461 1461 N -> S (IN REF. 2).
FT CONFLICT 1581 1581 G -> S (IN REF. 2).
FT CONFLICT 1600 1600 I -> V (IN REF. 2).
FT CONFLICT 1661 1661 R -> S (IN REF. 2).
FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

```

```

Query Match 4.1%; Score 160; DB 1; Length 1790;
Best Local Similarity 18.2%; Pred. No. 1.8;
Matches 142; Conservative 145; Mismatches 288; Indels 206; Gaps 32;
Qy 23 NLEVIOAEVKQENRLNSESQGLGYPFDLNFQAPVYTSSTGCD-----LSI 74
Db 1023 NIEQKKTISDLEQYKEIKESDSSKDEYQISLKEKLETTATANDENVNKISLTK 1082
Qy 75 PSSELE-----NIPSENYQFQSAIWSGFIKVKKSDY-----TPTASDNHYTM 118
Db 1083 TRELEAEIAAYKNKLNLEKLETKLESEKALKEVENEHLEKEIKLEKEATEFKQOLNS 1142
Qy 119 WVDQEVINKASNSKIKLEGRLOIKI-----QY-----GRENPTEGL 159
Db 1143 LRANLESEKHEHDLAOLK---YEEQIANKERYNEISQLNDELITSTQDESKTKK 1199
Qy 160 DFKIYW-----TSQKKKEVYSSDNLQPLKQSSNSRKRSTGAPTVPRDN 209
Db 1200 NDELEGEVKAKKSTSEDSNLSKSEIDALNLQIKELKKN-----ETNE 1243

```


QY 466 NOFELEKTKOL-----RLDTPDYVGN--IATYFENGVRVDTGSNMSEVLPOIQ 514
DB 1341 YKFLKEKRDFFLSYNYIKSDIDPINFADVGLGYILSEKYSDDSIKKYINDOG 1400
QY 515 ETTARIIFNGDNLVERRIAANVPSDLETTKPDMLKEALKIAFGNEPNGNLQYOGK 574
DB 1401 ENEKYLPF-----LNNIELLYKTVN-----DKIDLFFV-----IHLFAK 1433
QY 575 DITEFDFNFDOOTSONIKNOELNATNIYVLDKIKLNAMN--ILIRDRFHYDRNN- 631
DB 1434 VL-----NYTYEKS-NVEVKIKELN--YLKTIQDKLADFKKNNNFVGJADLSTDYNNHNL 1485
QY 632 ----IANG-ADESVYKHAHREVINSSTGGLLNTDKD--IKRILSGYVELEDTEGLKEV 684
DB 1486 LTKFLSTGMVFNELAKTVLSMLDGNLOG-MLNISOHQCVK-----QCPONSGCFRRH 1537
QY 685 INDRYDMLNLSLRDGGKTFIDFKKYNDKLPYISNPYKVNVAVTKENTLINSENG- 743
DB 1538 LDE-----RECCCLNATYKQEGDKC---VENPNTCN-----EN-----NGG 1571
QY 744 -----DTSNGIKRI 753
DB 1572 CDADAKTEEDSGSNG-KKI 1590
RESULT 13
MSPI_PLAFW ID MSPI_PLAFW STANDARD: PRT: 1639 AA.
AC P04933;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE MEZOZOITE SURFACE PROTEIN 1 PRECURSOR (MEZOZOITE SURFACE ANTIGENS)
DE (PMSA) (P195).
GN MSP-1.
OS Plasmodium falciparum (isolate Wellcome).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5648;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=6601435; PubMed=2995820;
RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
RA Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
RA Freeman R.R.;
RT "Primary structure of the precursor to the three major surface
RT antigens of Plasmodium falciparum mezozoites";
RL Nucleic Acids Res. 13:270-273(1985).
RN [2]
RP REVISIONS.
RA Holder A.A.;
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -1- PPM: MEZOZOITE SURFACE ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEZOZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: X02919; CAA26676.1; -
CC PIR: A24594; A24594.
CC InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF 1.
KW Malaria; Mezozoite; Polypeptide; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1639 MEZOZOITE SURFACE PROTEIN 1.

FT CARBOHYD 116 116 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 764 764 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 768 768 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 783 783 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 920 920 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 964 964 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 1058 1058 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 1174 1174 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 1445 1445 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 1526 1526 N-LINKED (GLCNAC . . .) (POTENTIAL).
SQ SEQUENCE 1639 AA; 187618 MW; 2C255B616C87F6E CRC64;
Query Match 4.1%; Score 159.5; DB 1; Length 1639;
Best Local Similarity 19.2%; Pred. No. 1.7;
Matches 165; Conservative 136; Mismatches 308; Indels 251; Gaps 42;
QY 23 NLEVIOAEVKQENRLNSESSESSOGLGYFSDNLFQAPMVYVSTTGDLSI--PSELE 80
DB 862 NLRYKLKNDKIKN-LLEAKKVTSTSVKTLSSSMQ---PLSLTPQDKREVGANDTSHST 917
QY 81 NIPSENOYFOSAINSG-----FIKKKSDERTPATSDNHYTMVDD 122
DB 918 NLNNSKLFEENILSLGKNNKIYQELIGKSSSENFYEKILKSDPFYNESTFNFKSKADD 977
QY 123 -----QEVINKASNS-----NKIRPLEKGRLY-----QIKIQVORENP 154
DB 978 INSLNDESKRRKLEEDIKLKKTLQSLFSDLNKYLKLERLFDKKKYKQKQKTKTL 1037
QY 155 TEKGLDFKLYWTDSONKKEVVISDNQLPELKOKSSNRKRSTAGPTVPDRNDGIPD 214
DB 1038 LKEQESKLT--NSLNPKHVL--QNFVSFNNKKRAELAELENT----- 1077
QY 215 SLEVBGYVDVKNKRTFLSPMISNIHEKKGLTKY---KSP-----EKWASTADPSDFE 266
DB 1078 -----LENTKILKHY-----KGLVYKNGSESPKLTLSERISIQTEQVLSLE 1120
QY 267 -----KVYGRIDKNVSPPEAR-----HPVAAVYIVHVDENILSKN---EDQST 308
DB 1121 NPKVLSKLEGLKIDNLNLEKKKLSYLSGLHHLA-----ELKEVYKNNKNGNSPSE 1173
QY 309 QNTD-----SETRTISKNTSTSRHTSEVHGNAAEVHANTSTRTSTSEVH 353
DB 1174 NNTDVNNALESYKFKLPBGTVATVAVSSESDTLQESQPKPKAPSYHVG-AASNTITTSQN 1232
QY 354 GNAEYHNAI-----DHSLSLACERTMAETMGLNTADTARLANANTYVTGTAPIYV 406
DB 1233 VDDEVDVILVIFGESEEDYDGLQVVTGAVTPSVYIDNLSKIENEYVLYKPLAGV 1292
QY 407 LPTTSLVYKGNQOTLTIKAKENQSLIAPNNVPSKMLAPLAAQDF--SSTPIPTMY 465
DB 1293 YRSLLKQLENNVMPFNVAVKILNSRFKRENF---KWLSEDLIPYKDLTSSNIVYKDP 1349
QY 466 NOFELEKTKOL-----RLDTPDYVGN--IATYFENGVRVDTGSNMSEVLPOIQ 514
DB 1350 YKFLKEKRDFFLSYNYIKSDIDPINFADVGLGYILSEKYSDDSIKKYINDOG 1409
QY 515 ETTARIIFNGDNLVERRIAANVPSDLETTKPDMLKEALKIAFGNEPNGNLQYOGK 574
DB 1410 ENEKYLPF-----LNNIELLYKTVN-----DKIDLFFV-----IHLFAK 1442
QY 575 DITEFDFNFDOOTSONIKNOELNATNIYVLDKIKLNAMN--ILIRDRFHYDRNN- 631
DB 1443 VL-----NYTYEKS-NVEVKIKELN--YLKTIQDKLADFKKNNNFVGJADLSTDYNNHNL 1494
QY 632 ----IANG-ADESVYKHAHREVINSSTGGLLNTDKD--IKRILSGYVELEDTEGLKEV 684
DB 1495 LTKFLSTGMVFNELAKTVLSMLDGNLOG-MLNISOHQCVK-----QCPONSGCFRRH 1546
QY 685 INDRYDMLNLSLRDGGKTFIDFKKYNDKLPYISNPYKVNVAVTKENTLINSENG- 743

```
Db 1547 LDE-----RECKCLLVNKGCGDKC---VENPNTCN-----EN-----NGG 1580
OY 744 -----DSTNGICKI 753
Db 1581 CDADAKCTEEDSGSNG-KKI 1599

RESULT 14
MDS3_YEAST STANDARD: PRT: 1487 AA.
ID MDS3_YEAST
AC P53094;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MDS3 PROTEIN (MCK1 DOSAGE SUPPRESSOR 3).
GN MDS3 OR YGL197M OR G1307.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;

RN SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=97197971; PubMed=9046087;
RA Coglieva M., Klima R., Bertani I., Delneri D., Zaccaria P.,
RA Bruschi C.V.;
RT "Sequencing of a 40.5 kb fragment located on the left arm of
RT chromosome VII from Saccharomyces cerevisiae."
RL Yeast 13:55-64(1997).
RN [2]
RP CHARACTERIZATION.
RA Li M.B., Neigeborn L.;
RL Unpublished observations (XXX-1997).
CC -I- FUNCTION: NOT KNOWN; NEGATIVE REGULATOR OF EARLY MEIOTIC GENE
CC EXPRESSION.
CC -I- SIMILARITY: TO YEAST YER132C.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X91837; CAA62947.1;
DR EMBL: Z72719; CAA96909.1;
DR SGD: S0003165; MDS3.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF01344; Kelch; 2.
KM Melosis.
SQ SEQUENCE 1487 AA; 167073 MW; 768AEFBAB796E447 CRC64;

Query Match 4.1%; Score 158.5; DB 1; Length 1487;
Best Local Similarity 19.1%; Pred. No. 1.7; indels 255; Gaps 42;
Matches 170; Conservative 145; Mismatches 318;

OY 14 STILVSTGNLEVIOAEVQENRLNESSSQGLGYFSPDNPQAPMVTVSS----- 67
Db 593 STILFSPHSSRNSSKAVQEGRL-----SSGSLDNVF-EKNF--PIFARTSVSEAON 643
OY 68 TTGDLSIPSELENTIPSENOYFOSAIWSGFIVKKSDEYTFATSDNHVTMMV----- 120
Db 644 TQPOVANADAKAPNTPS-----TSDDESSSSSDLYSTPHYQRNDDE 686
OY 121 DDOEVYN--KASNSKRIEKGRLYO--IKIQORENPPEKGLDFKLYWTS--ONKKE 173
Db 687 DDEDPVSPKPVSKNSI-----YPIRKTESSSTSSNGMIFVPEKEKAVTSNTE 739
OY 174 VTSNDLQIPELKOKSSN-----SRKRSTAGPTVPDRDNDGP-- 213
```

```
Db 740 ALLESNLSIQELSRRRSLKSIIPSGELLRSSISEAEHORASHPLTSSPLFEDSGTPCGK 799
OY 214 DSLVEGYIVDVNKKRTFLSP-----WISNIEKKLTYYKSSPEKMSASDP 261
Db 800 QLOOLOOHT--IONPHNHLSPPRFRSRSARSSISYVSSSDRRG-----NSISSRT-SDS 851
OY 262 YSDFEKVYGRIDKNVSPEARHP-----
Db 852 FGT-PPVLGVNLVPLPPQTRPEPPPPCPAMSTGSTRSNTLTDMMSNKAAPSSRR 910
OY 284 -----LVAAPYIVHMENTIISKNEQSTQNTDSETRTSKNTSTSRTH 328
Db 911 SSHIGRRSSPTENNAFASAPBRASLDGQMGKSLKSGSTQYQPRKNSPKANETIQTP 970
OY 329 TSEVHGNAEYHANTSTSRHT-SEVHGN-AEYHAVAIDHSLSL-----AGERMAETMGL 381
Db 971 TS--SNNEWSQSVTSNDSFDSLOSNFALDEPLLPTRSLVMPMPTSTVRAFAEFEPY 1027
OY 382 NTADTARLANINRYVMTGTAPIYVNLPTTSLVGNKQTLATIKAKENQSLQILAPNNYYP 441
Db 1028 GQVNSKWLAPVALDLVMAKITEIPLYKLIL--EVLSTLAKKESLSICTSLMET 1084
OY 442 SKNLAPIALNAODFSSTPTM--NNQPLELEKT-----KQLRLTDQV 484
Db 1085 FRRTKLINSYKGEDEKTNVTYLSNDNYQELKLKVSLENIDNGYVDPDLRKQSAOSSST 1144
OY 485 YGNIAIYFNENGRVYDTGSGNWSSEVLFOIETTARIIFNG-----KDLNL 529
Db 1145 QESGSGANGE--KTAGAGSLET--SSTWVPYFAGGPRDSHNSVSGISGFPNPMNI 1196
OY 530 VERRIAAVNPDPLETTKPPMTLKEALKIAFG--FNEPKNLOYGKDDTERDFENDDQ 586
Db 1197 QGSRRTSGFS-----PRVKMKSSLSKEIDPKTFYE--EYEPKESGPDNDQDQ 1244
OY 587 TSONIKR-QLAELN-----ATNIYT-----VLDRIKLNAKNNILIRDKRFH 626
Db 1245 TNGSFRHLHFDNNYSGISSSTNSISSDLEKEBEQLODLIEREDSAELIDARF- 1303
OY 627 YDRNNIAVGADSVYEAHREVINSSTEGLLINDIKRILSGYIEIDTEGLKEVIN 686
Db 1304 --RNK-----EDDVTXKD-----ISNDKKNRYLPHEKNNLKAKG--KETRDVREEEEDF 1350
OY 687 DRDMLNSSLRODGKTFIDFKKYNDKL-PLXISN--PNKKVNVYANT 731
Db 1351 FGLGMLSLNKKIKREAK-HVD--KYDQSVDPLEKSSAFPOSPIRAYGST 1395

RESULT 15
FAT1_SCHPO STANDARD: PRT: 1385 AA.
ID FAT1_SCHPO
AC O13735; Q90UTJ0;
DT 15-JUL-1998 (Rel. 36, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ACTIN INTERACTING PROTEIN 3 HOMOLOG.
GN FAT1 OR SPAC15A10.15 OR SPAC15E1.01.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RN SEQUENCE OF 1-1033 FROM N.A.
RC STRAIN=972;
RA Murphy L., Harris D., Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 1023-1385 FROM N.A.
RC STRAIN=972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Lyne M.H.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RN IDENTIFICATION, AND GENE NAME.
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:51:05 ; Search time 72.22 Seconds
(without alignments)
805.834 Million cell updates/sec

Title: US-09-747-521-4

Perfect score: 3913

Sequence: 1 MKRRKVLIPMALSTILVSS.....TSTNGIKKILFSSKKGYEIG 764

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR68:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3800	97.1	764	2	I39934 protective antigen
2	883.5	22.6	875	2	I40862 Iota toxin compone
3	235.5	6.0	192	2	I39933 cryptic protein -
4	235.5	6.0	204	2	G59104 hypothetical prote
5	205.5	5.3	4688	2	F82885 hypothetical prote
6	193.5	4.9	4919	2	T31105 hypothetical prote
7	193	4.9	1072	2	A86827 hypothetical prote
8	190	4.9	1302	1	JC6009 surface-located me
9	186.5	4.8	2401	2	T28676 rhoptry protein -
10	186	4.8	2178	2	S55805 alpha-toxin - Clos
11	183	4.7	1465	2	S31262 TYB protein - yeas
12	183	4.7	1803	2	S56894 TYB protein - yeas
13	182.5	4.7	1193	2	S68218 botulinum neurotox
14	182	4.7	1658	2	S55101 hypothetical prote
15	180	4.6	1051	2	T18351 ImpI protein - Myc
16	180	4.6	2269	2	T28677 rhoptry protein -
17	179.5	4.6	1365	2	T30822 ImpI protein - Myc
18	177.5	4.5	1467	2	PC1253 TYB protein - yeas
19	177	4.5	1315	2	T28679 fibrinogen-binding
20	177	4.5	1837	2	T41023 probable nuclear p
21	177	4.5	3724	2	T18427 hypothetical prote
22	176	4.5	1620	2	S61535 nucleotide-binding
23	175.5	4.5	1516	2	E71619 RAD2 endonuclease
24	175	4.5	1802	2	S52611 TYB protein - yeas
25	175	4.5	1939	2	T18372 repeat organellar
26	174	4.4	2529	2	B64635 toxin-like outer m
27	173.5	4.4	1230	2	S56850 SMCI protein homol
28	173.5	4.4	1271	2	D64237 hypothetical prote
29	173	4.4	1230	2	E64664 outer membrane pro

30	171.5	4.4	821	2	S67087 hypothetical prote
31	168.5	4.3	1115	2	T41342 probable coiled-co
32	168.5	4.3	2526	2	T20531 hypothetical prote
33	168	4.3	2523	2	T18477 hypothetical prote
34	168	4.3	4152	2	T31102 filamentous hemagg
35	167.5	4.3	786	2	T18469 ORF MSV156 hypothe
36	167	4.3	1127	2	T28317 surface membrane p
37	166.5	4.3	624	2	PC6003 myosin heavy chain
38	166.5	4.3	1928	2	S46773 hypothetical prote
39	165.5	4.2	1553	2	T18502 fibrinogen-binding
40	165	4.2	1166	2	T28680 hypothetical prote
41	165	4.2	5005	2	F82884 p101 protein precu
42	163.5	4.2	888	2	A38539 protein with DnaJ
43	163.5	4.2	1014	2	H71602 toxin-like outer m
44	163.5	4.2	2399	2	H71879 cell surface antic
45	162.5	4.2	2340	2	B71704

ALIGNMENTS

```
RESULT
1
I39934
protective antigen precursor - Bacillus anthracis plasmid
C:Species: Bacillus anthracis
C:Date: 19-Jul-1996 #sequence, revision 19-Jul-1996 #text_change 01-Dec-2000
C:Accession: I39934; S69160; F59104
R:Welkos, S.L.; Lowe, J.R.; Eden-McCutchan, F.; Vodkin, M.; Leppla, S.H.; Schmidt, J.
Gene 69, 287-300, 1988
A:Title: Sequence and analysis of the DNA encoding protective antigen of Bacillus ant
A:Reference number: I39933; MUID:99172073
A:Accession: I39934
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-764 <RES>
A:Cross-references: GB:M22589; NID:g413280; PIDN:AAA22637.1; PID:g413282
R:Friedman, T.C.; Gordon, V.M.; Leppla, S.H.; Kimpel, K.R.; Birch, N.P.; Loth, Y.P.
Arch. Biochem. Biophys. 316, 5-13, 1995
A:Title: In vitro processing of anthrax toxin protective antigen by recombinant PC1 (
A:Reference number: S69160; MUID:95142670
A:Accession: S69160
A:Molecule type: protein
A:Residues: 197-202 <FR>
R:Okinka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koeh
J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of pXOI, the large Bacillus anthracis plasmid harb
A:Reference number: A59091; MUID:99445483
A:Accession: F59104
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-313, 'O', 315-764 <OKT>
A:Cross-references: GB:AF065404; NID:g4894216; PIDN:AAD32414.1; PID:g4894326
A:Experimental source: strain Sterne
A>Note: similar to anthrax toxin moiety, protective antigen, paga formerly pag, plasm
C:Genetics:
A:Gene: pXOI-110
A:Genome: Plasmid
C:Function:
A:Description: three component exotoxin; protective antigen binds to receptors on the
y active components edema factor or lethal factor; the complex is internalized by rec
C:Keywords: exotoxin
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-196/Domain: propeptide #status predicted <PRO>
F:197-202/Product: protective antigen #status experimental <MAT>
```

Query Match 97.1% Score 3800; DB 2; Length 764;
Best Local Similarity 97.6% Pred. No. 8.1e-186;
Matches 746; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

Qy 1 MKRRKVLIPMALSTILVSSGTNGEIVIOAEVKKOENRLNSESSESSQGLGYFFDLNFOA 60
Db 1 MKRRKVLIPMALSTILVSSGTNGEIVIOAEVKKOENRLNSESSESSQGLGYFFDLNFOA 60

QY	61	PMVYVSTTGDLSPISPESELENIPSENOYFQSAIWSGFIYVKKSDSEYTPMPSADNHTMAY	120
Db	61	PMVYVSTTGDLSPISPESELENIPSENOYFQSAIWSGFIYVKKSDSEYTPMPSADNHTMAY	120
QY	121	DDQEVYINKASNSNRIKLENGRLYQIKIYORENPTEKGLDFKLYWTDSONKKEVISSDNL	180
Db	121	DDQEVYINKASNSNRIKLENGRLYQIKIYORENPTEKGLDFKLYWTDSONKKEVISSDNL	180
QY	181	QIPELKOKSSNRKRKRSSTAGPVPDPRNDGIPDLSLEVGGYVYDVKNKRTFLSPMISNH	240
Db	181	QIPELKOKSSNRKRKRSSTAGPVPDPRNDGIPDLSLEVGGYVYDVKNKRTFLSPMISNH	240
QY	241	EKKGLTKYSSSEKMSSTADPYSDFEKYVGRIDKNVSPARHPVLAAYPIVHVDMENIL	300
Db	241	EKKGLTKYSSSEKMSSTADPYSDFEKYVGRIDKNVSPARHPVLAAYPIVHVDMENIL	300
QY	301	SKNEDOSTQNTDSEFTTISKNTSTSRTHSEVHGNAEVAHNTSTSRTHSEVHGNAEVA	360
Db	301	SKNEDOSTQNTDSEFTTISKNTSTSRTHSEVHGNAEVAHNTSTSRTHSEVHGNAEVA	360
QY	361	VAIDHSLSLAGERTAAETMGLTAOTARLANANRYVNTGTAPLYNLPTTSLYLGKNTL	420
Db	361	VAIDHSLSLAGERTAAETMGLTAOTARLANANRYVNTGTAPLYNLPTTSLYLGKNTL	420
QY	421	ATIKAKENOLSIILAPNNYPPSKNLAPIALNADDESSPTIMNYNOFLEKTRKOLRD	480
Db	421	ATIKAKENOLSIILAPNNYPPSKNLAPIALNADDESSPTIMNYNOFLEKTRKOLRD	480
QY	481	TDOYVGNIAITYNPENGRVADTGSNMSEVLPOIETTARIIFNGKDLNVERRIAVNPS	540
Db	481	TDOYVGNIAITYNPENGRVADTGSNMSEVLPOIETTARIIFNGKDLNVERRIAVNPS	540
QY	541	DPLETTKPMPTLKEALKIAPGFNEPNGNLQYOGKDIETEDFNPDQOSTONIKKQALNELA	600
Db	541	DPLETTKPMPTLKEALKIAPGFNEPNGNLQYOGKDIETEDFNPDQOSTONIKKQALNELA	600
QY	601	TNIYTVLDKIKLNKANNILIRDKRFHYDRNNIYAVGADSEVYEAHREVIINSTEGLLNI	660
Db	601	TNIYTVLDKIKLNKANNILIRDKRFHYDRNNIYAVGADSEVYEAHREVIINSTEGLLNI	660
QY	661	DKDIRKILSGYIYEIEDTGLKEVINDRYDMLNSSLRODGKTFIDPKKYNDKPLIYISN	720
Db	661	DKDIRKILSGYIYEIEDTGLKEVINDRYDMLNSSLRODGKTFIDPKKYNDKPLIYISN	720
QY	721	PMYKVVVYAVTKENTTIINSENGDSTNGIKKILIPSKGYEIG	764
Db	721	PMYKVVVYAVTKENTTIINSENGDSTNGIKKILIPSKGYEIG	764
RESULT	2		
	140862		
	lota toxin component Ib - Clostridium perfringens		
	C:Species: Clostridium perfringens		
	C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 15-Oct-1999		
	C:Accession: 140862; 542774		
	R:Perelle, S.; Gilbert, M.; Bogueff, P.; Popoff, M. R.		
	Infect. Immun. 61, 5147-5156, 1993		
	A:Title: Characterization of Clostridium perfringens lota-toxin genes and expression in		
	A:Reference number: 140861; MUID:94041637		
	A:Accession: 140862		
	A:Status: preliminary; translated from GB/EMBL/DBJ		
	A:Molecule type: DNA		
	A:Residues: 1-875 <RES>		
	A:Cross-references: EMBL:X73562; NID:g929031; PION:CA51960.1; PID:g414655		
	Query Match		
	Best Local Similarity 31.5%; Pred. No. 1,8e-37;		
	Matches 262; Conservative 135; Mismatches 279; Indels 155; Gaps 35;		
QY	15	TLIVSTGLEY-----IOAEVKQENRLINESSSOGILGYIFSDLNFOAPMVYVTSSTT	69

Query Match 6.0% Score 235.5; DB 2; Length 192;

Db 15 TAMIQSITISYNYAQTTONQNDTNOKEEITNEUTLSSNGMGVYFADEHFKDLELMAPIKN 74
16 GDLSPSSSELENIPBE-NQYFOSAIWGSFIAKKKSDTEFTATSNADNHVTAMVDQEVLYNK 128
17 75 GDLFEEKKYVDLLTEDNSSIKSIRMTGIIPSEDGEIYLSTDR-NDVLMOINKAGDIAK 133
18 129 ASNSKKRIELEKRLQIKROYOREN-----PTEKGLDFKLYWTDSOKKVEISSDN 179
19 134 ---TLKVNMKKQAYINRIEIDKKMLGSLDLSVP-----KLIV-ELNGKTIYIIPEN 182
Db 180 LQLPELKOKSSNRKKRSTSAQTPV-----DRDNDGIPDSLEYEG 220
21 183 LFEPDYSKIDEND-----PFIPNNNFDPVFREFSAMEDEDLDTDNDNIIPAYEKNG 233
Db 221 YTVDAKNKKTETLSPISINHEKKGLTKSSPEKASTADPSDEKATGDIAKNVSEPA 280
22 234 YTI---KDSIAVKMNDSEFAE-QGYKKYVSYSLESNTAGDDPYTDQKASGISDKAIKLEA 288
Db 281 RHPLAAAPYIVHDMENIIISKNEDOSQNDSERTISAKTSTSRTHTSEVHGNAEHA 340
23 289 RDLPLAAAPYVGVGEMENLISTNEHASS---DOGRITYSRATTSKTDANTV--GVSTISA 342
Db 341 -----NTSRPTSEVHGNAEVAHVLAVIDISLAGEPTMAETMWGLANTADARLAN 392
25 343 GYONGFTGITTYSHTTD-----NSTAVODS---NGE-SMNGLSINNKGESYINAN 391
Db 393 IRYVNTGTAPIYNVLPTTSLVLGKNOTLATITAKKENOLSQLAPNNYYPSKNLAPIALNA 452
27 392 VRYVNTGTAPMYKMPPTNLVLT-DGETLATIKAONQIGNNLSPRETYPKRGSLPALNT 450
Db 453 QPDFSPTPMNNGOFELEKTKOJLRDTDOYGGINATFYFNENGVRARDTSNNSEVLPO 512
29 451 MOGFARLPINWDOLKILDSGQKILETQVSGVGTRN-SQGCI-TTEENSMSNYSIQ 508
Db 513 IOETARIIFNGKDLNLVERARIAAVNPDELTTYPMDTLKEALKIAGFEDBPNGLOY- 571
31 509 IDSVASAITLD-TGSOTFERRAVAKEQGNPEKTI-PETTIGEAIKKANSAIK-NGELLTF 565
Db 572 OCKDITE--FDENPDQOTSQNIKNOJLABLNATINYTLDKIKLANKNILIRDKRF-HY 627
33 566 NCIPIDESCEVELIFPDNTNSEIIEKQKLYLDOKKIYVN---KLERGMNILIKVSYETFN 621
Db 628 DR-NNIIVAGADSIVKKAHREVINSTGL-----LNIIDKDIKILISGY----- 671
35 622 DEYNNEP--ASMSNIDTKMQDGLQSVANKLSETKIIIPMSKLPKRYKRVSGYSKODPT 679
Db 672 ---IVEIEDTGLEKVINIDRYDMLNIS-----SLROQKFIFPKKND 712
37 680 SNTSVNIKSKQCKTDYLVPEKDYTKFSEYEFTTKDSSDIETILTSSGVFLDNLSITE 739
Db 713 --KLPLYISNPYKY-----NVYAATKENTIIINPSENCDTSTNGI 750
39 740 LNSTPEILKEPEIKVPDQELDAHNNKYADIKLDT-----NTGMTYIDGI 785

RESULT 3
339933
Cryptic protein - Bacillus anthracis
C.Species: Bacillus anthracis
C.Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 15-Oct-1999
C.Accession: J39933
R.Welkos, S.L.; Lowe, J.R.; Eden-McCutchan, F.; Vodkin, M.; Leppla, S.H.; Schmidt, J.
Gene 69, 287-300, 1988
A.Title: Sequence and analysis of the DNA encoding protective antigen of Bacillus anthracis
A.Reference number: J39933; MUID:89172073
A.Accession: J39933
A.Status: preliminary; translated from GB/EMBL/DBU
A.Molecule type: DNA
A.Residues: 1-192 <RES>
A.Cross-references: GB:M22589; NID:g143280; PIDN:AAA22636.1; PID:g143281

[illegible][illegible]

[illegible]

```

OY 396 VITGAPLYNVLPSTSLVGNKQTLATIK---KENLSQLAPNNYPSKNLAPIALNA 452
Db 810 ANTLAKLTD-----KDNITQQAQTELEKEVQCANAVASNNTASQSAKSSLDA 859
OY 453 Q-DDFSPTITMWNQFLELEKTKQLRLDQVYGNIAATYVFNENGRRVDTGSMSEVLP 511
Db 860 KVTETTKLETFENKQKDKVFEKELEQTRKDIDF-----INTNKNPDPST 904
OY 512 QIQEFTALINGKQLNLVERRIAAVNPSPDELTTPKDMTLKEALKIAFGNENGNLQY 571
Db 905 LISELTSK-----RDSKNSTTSSNSDSLETATTELQKALAKA----- 942
OY 572 QGKDTTFEDFNDOQTSNIKQLAELNATNYIVYLKIKINAKMNILIRDKRFHYRNN 631
Db 943 -NTDKDQAD-NLARSTKQLNKSISAN-----FLAKLT-----DKDN 979
OY 632 IAVGAD---ESVYKAAHAEVINSSTEGLLINDKDIRILSGIYEIEDTGLKEVINDR 688
Db 980 TIQAQTELEKEVQCANAVASNNNTASM-----OSAKSSLDAKYETI--TKLETFPNKD 1032
OY 689 -YDMANISLRDQGTKEFIDFKFYNDKLEPLYSNPYKVVAVT---KENTINPSNG 743
Db 1033 DYKFELEQTRKDIDFINTNK-----TNPNYSLIELTSKRSKSNSTNSSKS 1083
OY 744 DTST 747
Db 1084 DIET 1087

RESULT 9
T28676
rhoptry protein - Plasmodium yoelii (fragment)
C.Species: Plasmodium yoelii
C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C.Accession: T28676; A:45521
R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A.Title: Comparison of two members of a multigene family coding for high-molecular ma
A.Reference number: 220507; MUID:97077455
A.Accession: T28676
A.Status: preliminary; translated from GR/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1*2401 <SIN>
A.Cross-references: EMBL:U36927; NID:g10411784; PID:g10411785; PIDN:AA841263.1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A.Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple
A.Reference number: A45521; MUID:91101660
A.Accession: A45521
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 2260-2401 <KE>
A.Cross-references: GB:M34281

Query Match 4.8%; Score 186.5; DB 2; Length 2401;
Best Local Similarity 21.7%; Pred. No. 0.21; Mismatches 313; Indels 219; Gaps 44;
Matches 182; Conservative 123;

OY 30 EVKQENRLNEESSQCLAGYFSDLN-----EQAPVAVTSSTTGDSL I 74
Db 402 EVAKENQVLNVKKSILIEKKHYNDQIMINDIKKEKAKQNDQREKHKITPRNEMKYOK 461
OY 75 PSELENIPEENQYFQSAI--WSGFIKY--KSDYTFATSA DNHVTYTMVDOEV---I 126
Db 462 PSIEIKIKMDE---FLSKVNVKYNDFDKYKKEVESEHKNFTELTNKITEVSDSEIKYE 518
OY 127 NKASNNKIRLEKGLYQIKIOYQRENFTEGDLGFKLWTPSQKKKEVSISSNDLOPELK 186
Db 519 NKFNDQSKLINETKK--SIEEYONINTLKKVDD---YIKVCLNTNELITCNKKKOTTLK 573
OY 187 OKSSNSRK--KRSTSGAGTVPDRDNDGIPD-----SLEEGYTVDVAKNRKRTLSP 234

```

```
Db 574 DKLNONIKITIKETNSIDKIYTKQFENILLDKTELETKTGLSLNNHESNNKELLTYFD 633
Oy 235 WISNIHEKKGLTKYKSSPEKMWSTASDPYSDFEKVTGRIDKNVSPARHPLVAAPYIVHD 294
Db 634 LKANIGKKNENMLYKOFNEK-----EKAVEDIKKKNDIKNIVSN-----IEIT 677
Oy 295 MNILLSKNEDOSTONTDEFTTISKNTSTSTHTSEVGNAEVANTSTSTHTSEVHG 354
Db 678 IYTSYINNEDE-----TERE-----IGKSIELLNTKYLE-----KXKANV-----TNNIEIKE 720
Oy 355 NAEVAVAIDHSLAGERTMAETWGLNTADPARLANIRYVNTGAPRYNLTPTSTLV 414
Db 721 KKK-----DYFODEPK-----EKNIKYDEN-----KKNIIDILNOKI 755
Oy 415 GKN-OTLATIKA-KENQUSOI-----LAPN-----NYRP-----SKNLAPIA----- 449
Db 756 DKSIETLEIKKSNENHIDEIKQIDKLKKVKNKTMFNEDPKIEKKIENIYEKIDKKKN 815
Oy 450 LNAODDFSTPTMNYNQFLEKTKQLRDPDYVGNATYATNFENGRRVDTGSMSEV 509
Db 816 IYKEIDKLNEISKIENKTSLEKLKNINLSYKSLGNLFLOQIDBEKKKAHTTKAMEA 875
Oy 510 ----LPOIETTARIIFENGKDLN-----VERRIAVNPSPD-----LETT-----KP 548
Db 876 YIIDLNDIKKKKQOI-----EKEMINMDIKMDIHEKMAKALINSHDQKIYHTSKNHEKI 932
Oy 549 DMTLEKALKIAFGNEP-----NGNILOYQGDITEFDFNPDQSTONIKQOLAEAFNTIYT 605
Db 933 SDIRKNSLIIDPFSEESYINDIKKKELEKNVLE-----SQNNNTDINOGLSKIE--NIYN 985
Oy 606 VLDKIKLAKKNILIRDRF--HYDRNNIAGADESVYEAHREVIINSTBEILLNIDKD 663
Db 986 IL--KLNKIKKIKDKVXEYDEIEKNN-----KKIAELSNS----- 1020
Oy 664 IRKILSGYIEIETDEGLKE-----VINDRY-----DMINISLRDQGTFFDKKYNDK 713
Db 1021 -EKI-----ITQKENSLSKEQSKIKSTIDNYVSECIKNIINL--KTYIVNEKN-- 1069
Oy 714 LPLYSN--PNKYVNV-----VAVTKENTIIINSENGDTSFN--GIKKILIEFSK 759
Db 1070 INTYFKNAEYVNONVSLNFENNEMADTFKSQYILTIKKNGTNNYDNIKELKEHKKK 1126

RESULT 10
alpha-toxin - Clostridium novyi (ATCC 19402)
C:Species: Clostridium novyi
A:Variety: ATCC 19402
C:Date: 28-Oct-1996 #sequence revision 08-Nov-1996 #text change 05-Nov-1999
R:Holmann, F.; Herrmann, A.; Habermann, E.; von Eichel-Streiber, C.
Mol. Gen. Genet. 247, 670-679, 1995
A:Title: Sequencing and analysis of the gene encoding the alpha-toxin of Clostridium novyi
A:Reference number: 140834; MUID:95342160
A:Accession: S55805
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-2178 <HOF>
A:Cross-references: EMBL:Z48636; NID:9728537; PIDD:CA488565.1; PID:9755724
A:Accession: S711294
A:Molecule type: protein
A:Residues: 1-15 <HOF>
R:Holmann, F.
submitted to the EMBL Data Library, March 1995
A:Reference number: S71158
A:Accession: S71158
A:Molecule type: DNA
A:Residues: 1-1179, 'LKV', 1183, 'LVTHIGE', 1191-2178 <HOS>
A:Cross-references: EMBL:Z48636; NID:9728537; PIDD:CA488565.1; PID:9755724
R:Holmann, F.; Habermann, E.; von Eichel-Streiber, C.
submitted to the EMBL Data Library, July 1993
A:Description: Sequence analysis of Clostridium novyi alpha-toxin: a member of the family
```

```
A:Accession: S44273
A:Molecule type: DNA
A:Residues: 1-243;1204-2178 <HOF>
A:Cross-references: EMBL:Z23281
C:Genetics:
A:Gene: tcn-alpha
C:Superfamily: cpl repeat homology
C:Keywords: virulence factor
F:1880-1899/Domain: cpl repeat homology <CP12>

Query Match 4.8%; Score 186; DB 2; Length 2178;
Best Local Similarity 20.8%; Pred. No. 0.19;
Matches 165; Conservative 116; Mismatches 264; Indels 250; Gaps 39;

Oy 122 DOEVIKASNSKIRLEKGLYQI-KIYOIREN-----PREKGLDFL----- 163
Db 202 DIIINFLSNRYK-----YDIGKLNQKKNKMAIGATIDINTENILTNKLKSY 253
Oy 164 YWTDSONKKEVISSDNLQPELKQSSNSRKRSTAGPTVPDRNDGIPDLE-VEGYT 222
Db 254 YQELIOTNNLAASDILRIALIKKYG-----GYCOLDPLPGVN 292
Oy 223 VDVKKKRFPLSPWISN-----IHEKGLTKYKSSPEKMWSTASDPYSDFEKVTGRI 272
Db 293 LSLFNDISKPNQMSDNYWEALFEALIANEKKLMNY-----PKYMEQVPEI 340
Oy 273 DKNVSPPEAR-----HPLVAAPYIVHDMEIILSKNEDOSTONTDEFTTISKNTS-TSRT 327
Db 341 KERILSFVRNHINDLILPLGDIKISOLEILSLKAAKTGKTFENAFIISNDSLTINN 400
Oy 328 HTSEVHG-----NAEVHANTSTSRHTSEVGNAEVAHAIDHSLAGERTMAETWG-L 381
Db 401 LISQLENRYEILINSIIQEFKICETFYDSYINSVSELYETTFPKNLSMGSSSFYQIIGVL 460
Oy 382 NTADPARLANIRY-----VNTGTAPIYVNLPTT-SYLVGKNQOTLATIKAKENQSOLA 435
Db 461 SSGFRPEVNSIVFFSGPNYISSATCDTYHFKNTPDMISSQOET-----FEA 508
Oy 436 PNNYPSKNLAPIALNAODDFESPTI-----TNMNYQFLEKTKQLRLDPDYVGNATYVN 492
Db 509 SNMLYFSK-----THDEFKSSWLLNSNIAEKFEOLIKITYGR-----TLN 549
Oy 493 FENGR-----VRVDTSQNMSEVLPQIETTARIIFENGKDLNVERRIAVNPSPDLETTK 547
Db 550 YEDGILFNKMKRVTT-----SELLKVIIEVNSSTRKYENYDLMNI-----LQIOG 593
Oy 548 PMTLEKALKIAFGNEPNGNQYQGKDITEFDFNQ--QTSQNIKKQOLAEIN----- 599
Db 594 DDISYSAVNV-FGKN-PKKSILIOGVDDFANVFYFENGIVQSDNINILSFNDIKRIK 651
Oy 600 -----ATNIYTVLDKIKLN-----AKMNIL----- 619
Db 652 LTLIGHGENVFPKLEGGKTYVNDLTNNIKRQLHLLRECVILKNKYKLNIILGCMFT 711
Oy 620 -----IRDKRFH-----YDRNNIAGADESVY--KEAHREVINSTEGLL 658
Db 712 PKVDINSTFVGKLFENKISRDLQPKGFSKNOLEISANKYAIRINREGCKREVLDYFGK-AVS 770
Oy 659 NIDKDIRKILSGYIV--EIBDT-----EGLEKY-----INDRDMLNISLRDQ-DKT 703
Db 771 NTDLIABOISNKKYVYVWNEVENTLSARVEQLKKVAFPAKDNISIIQTTNNOELKSLVNT 830
Oy 704 FIDF-----KKYNDKLPLYISNPNKYVAVTKE-----NTIINSENGDTSYNG 749
Db 831 YADLITFLYSELKREDIFPELDNIQIKERI--ILNEISRHDHFSNIIIDPFQKNISNM 888
Oy 750 IKKIL-----IFSKKGY 761
Db 889 I--ILFDSIIIEKDY 901

RESULT 11
```

[illegible][illegible]

611 KLNAAKNIILRDKRPHYDRN-----NIANGADESVYKHAHRE-----VI----- 649
Db 1621 -----VVKDIGIHYDDCKKDKKKVIAITDASVGEYDADSRIGVILIYGMNIPNVYS 1673
Qy 650 NSTEGLLINIDKIRKISGY-----IIVETEDP-----EGKK----- 682
Db 1674 NKSTNACVSTAEHLAIEGYADSETLVLTKELEGEGDNNDIWMITDSKPAIOGLNRSY 1733
Qy 683 -----EVINDRYDMLNISSLRDQKTFI-----DFKKY 710
Db 1734 QOPKEFTWIKTEIKERIKESIKLKITGKGNINADLLTKPVSADFKRF 1784
RESULT 13
568218
botulinum neurotoxin type A, nontoxic/nonhemagglutinating - Clostridium botulinum (strain
A:Species: Clostridium botulinum
A:Variety: strain NIH
C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 15-Oct-1999
C:Accession: S68218; S74301
R:Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
PDBS Lett. 376, 41-44, 1995
A:Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components
A:Reference number: S67988; MUID:96096783
A:Accession: S68218
A:Molecule type: DNA
A:Residues: 1-1193 <FUJ>
A:Cross-references: EMBL:D67030; NID:g2160224; PIDN:BAAL1050.1; PID:d1011710; PID:g11324
A:Experimental source: strain NIH
A:Accession: S74301
A:Molecule type: protein
A:Residues: 1-13;145-155 <FUJ>
A:Experimental source: strain NIH
C:Genetics:
A:Gene: ant
C:Keywords: neurotoxin
F:1-1193/Product: botulinum neurotoxin type A, nontoxic/nonhemagglutinating #status expe

Query Match 4.7%; Score 182.5; DB 2; Length 1193;
Best Local Similarity 18.6%; Pred. No. 0.12;
Matches 187; Conservative 144; Mismatches 351; Indels 323; Gaps 47;
14 STILVSTGNELVIAEYKQENRLNSESSSOGL-----LGYFSDLNFOQPM 62
Db 167 SNIVIEFG-----ANIYENTVFYKKEDANGMGTMTEIWFQPLFYKYDEFYIDPAI 220
Qy 63 VYTS-----STGDLSTP-----SSELENIPEQYQSAIMWGFYKVK--KSPD 105
Db 221 ELIKCLISLVFLYGIKPSDDLVIPIRLRSELENIYSQLNIVDLLVSGIDPKFTINTDP 280
Qy 106 YTFATS-----ADNHTVMVADQEVINKASNSKIKLER-----GLYQIKIQY-- 149
Db 281 YWFTDNYNSNAKKVEDHNIYETELEGNNALGNDIKLRLKQKFRNTINDIMELNLYTS 340
Qy 150 -----QREPTKGLDFKLYW-----TDSQNKKE 173
Db 341 KEFSIMPDRFNNALKHFYKQYKIDYPENYSINGFVNGOINAOISLSDRMODIINKPE 400
Qy 174 VYSS-----DNQLP-----ELKOKSSNKKKSTSG 201
Db 401 EITINLNGNNVSLMNSNIYDGCLKSTVDDFYSNYKIPYRAVEYHFNNSDSDLNVNG 460
Qy 202 PTVPRDNDGIPDSLEVEGYVD-----VKNKRTPLS--PAISNIHEKKGLTKYKS 250
Db 461 VI-----DNIPETIDVMPYKENCOKFSPVOKITSTREINTIIPPIIN-----LOAONT 509
Qy 251 SPEKSTASDPYDEKVTGRIDKNVSPARHPVLAAYPIVHDMENIT--LSKNEDOST 308
Db 510 NNEKFSLS-----SDPEVEYVSSKDKS-----LVYSFLSNVWFYLDISKDNP 551
Qy 309 QNTDSE-----FTTISKNTSTSRTHTSEVHGNAEVAHANT-----STSRTHTSEVHG 354
552 IDTKKYYVLMLEIFRNFYSFDITATOEINTNGCINKVVTWFGKALNILETSDSFVEEFON 611
Qy 355 NAEVHVALIDHSLSLAGEPTWA---ETMGLNADTARLANANIRYVYG--TAPIVNVLP 409
Db 612 LGAISLJNKKENLSPIITSYELPNDMLGLPLNDLNEKLFNYSKNTAVFKKITYINFLDQ 671
Qy 410 -----TSLVICKNOTLA--TIKAKENOLSOILAPNNYPSKNLAPIALNA---- 452
Db 672 WMTQYYSQYFDLICMAKRSVLAOETLIKRIIOKKLSYLLGNSNISDNLALNLTPTNTL 731
Qy 453 ODDFSSPTITM-----YNOFLEKTKOLRUDTOYVG--NIATYN 492
Db 732 RDISNESQIAMNNVDSFLNMAAICVFESNIYKPFISF-----MEDCINNINIKTKE 782
Qy 493 FENGRRVYTGSNMSEVLPOIETTARIIFNGKDLNLVERIAAVNPDPLETPDMTL 552
Db 783 F-----IQCTNINDEKLOLINON-----VFNSLDEFLN-----IQNKSLFSSETALIT 829
Qy 553 KEAL-----KIAFGFNEPNCN-----LOYQKDI--TFEDPND-----OOTSON 590
Db 830 KEETWPELVLYAFKEPQNNVIGDASGKNTSIEY--SKDIGLYGINSDALYLGNSQGIS 888
Qy 591 INQOLAELNATNITYV-----LDKIKLNAM----- 616
Db 889 FSNDEFENGCTNSFSISYFWLRNMGDTIKSLKLSKEDNCGEIYFODTGLVFNHIDSNG 948
Qy 617 ---NILIRD---KRFHY-----DR--NNIANGADESVY--KEAHREYVSTEGLLINDK 662
Db 949 NEKNITLSVSNNSMHTYITSDRLKEQLLITIDDLVLANESIKELINITYSNITSLISE 1008
Qy 663 DIRKLISGYIEIEDTEGLKEVINDRYDMLNITSLRQDQKTEIDFKRYNDKLPYLSMPN 722
Db 1009 NNPSYIEGLTILNKPTTS--QEVLSNWFYEVLLNSYIRDSNEERLE---YKTYQLY-----N 1060
Qy 723 YKVN---YVAYTKENTI---INPSENGDTSTGKIKIILFSKGY 761
Db 1061 YVFSOKPICEVAKONNNIYTLTNTNMLNLQASKFKLISINPKQY 1105
RESULT 14
555101
hypothetical protein YMR219w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YMR261.13; hypothetical protein YMR959.01
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 09-Mar-1996 #text_change 29-Oct-1999
C:Accession: S55101; S57587
R:Bedman, K.; Brown, D.; Bowman, S.
submitted to the EMBL Data Library, June 1995
A:Reference number: S55089
A:Accession: S55101
A:Molecule type: DNA
A:Residues: 1-711 <DED>
A:Cross-references: EMBL:Z49809; NID:g854459; PIDN:CAA89934.1; PID:g854471; MIPS:YMR2
A:Experimental source: strain AB972
R:Skellton, J.; Churcher, C.M.
submitted to the EMBL Data Library, June 1995
A:Reference number: S57587
A:Accession: S57587
A:Molecule type: DNA
A:Residues: 608-1658 <SKF>
A:Cross-references: EMBL:Z49939; NID:g887599; PIDN:CAA90190.1; PID:g887600; MIPS:YMR2
A:Experimental source: strain AB972
A:Map position: 13R
Query Match 4.7%; Score 182; DB 2; Length 1658;
Best Local Similarity 19.3%; Pred. No. 0.21;
Matches 161; Conservative 133; Mismatches 308; Indels 224; Gaps 39;
25 EYIOAEVQENLNLNSES--SSOGLLGYYFSDLNFOQPMVYVTSSTGDLSTPSSLENIIP 83

Db 287 EGVMELEDDIDVESDAEKDESOGAEGTEHS -VDFSKYMQPTDNTKIPVIEKYESDEHK 345
QY 84 SEMQYQSAIWS-GFIKKVKSDE-----YFATSADN---HYTMVVDQVINKASN 131
Db 346 VHQRYSEDGAFDFGSVNIVSDDESEDESOAESYSAENAVYHNEHEDKELIEDIES 405
QY 132 SNKIRLEKGLYQIKIOYQENPTEKGLDFKLT-----WTDSQNK 171
Db 406 SDS-----ESQASQESQSEDEDFEYKMKNEKSTSETEENTSROGAKADYTK 436
QY 172 KEVISSDNLOLPEL-----KQSSNSRKKRSTAGPTVPDRDNGIPDLS 216
Db 457 NKVEQOEENDEEPEKDDIIRSSLDKNFHNKKEYSNPVLENETDPAIVERNO-IND-- 513
QY 217 EVEGYVVDVKNKRTFL--SPWISNIHE--KKGITKYSK--PEKWTASDPY---S 263
Db 514 -VEGYVVTCKSVESDLHEHSP--DNLYDLAARMLQFOQSRNSCPOKEQVSESYLCHS 570
QY 264 DFEKVTGR-IDKNVSPEARHPLVAAYPIVHVMENIILSKNEQOSTQNTDSERTISKNT 332
Db 571 NGSNLISGRIDES---EEOIPLKD-----FTGENNNLKTDRGDLSSVEI 613
QY 323 STSRRTSEVHGNAEVH-ANTSTSRHTSEVHGNAEVAIDHSLSLAGEERTWETMGL 381
Db 614 EYEVKSEKKIDGSTEKELVPLSTDTIINNSLGNEDSIYSLDDADAISENLDPVLEI 673
QY 382 NTA-----DTARLNMIRVYNTGTATVYVLPPTSILVGLKNQTLAT 422
Db 674 KTTPKTEVVISSESVSTSEYEDNTVAMPQVEY---TSPFMD-PENSL---ND--D 721
QY 423 IKAKENQLOIILAPNNYSPSKNAPIALNAQDDFSSPTIM-----NYNOFLEK 474
Db 722 YEKKHDLKSTLAA-----LAPAFKKDAFEVAGVTKSCLTISGHTNFHSKET 773
QY 475 KQRLDLDQVYGNIAVYNEENGHRVVDGSMSEVL-PQIOETTARILFNGKDLNVERRI 534
Db 774 KOVS-DLDESTEYVTEFENEWTG---DENKQSKNFGVANSSTDKSTEDWTD---EKYE 824
QY 535 AAVNPSPLETTKPDMTLEKALKIAGFNEPNNGNLOYOGDITEPDPNF-DQOTSQ-- 590
Db 825 SAIVNYN---VHGSSCEBIIETASNVEE---NLRYCEDMNEAESQSDDEVCKDDG 877
QY 591 IKQOLA-----ELNATNIYVLDKIKLAKAKNIIIRDKRFYDRNNIIVAGDESVKE 643
Db 878 SKTQISFSDSPNQESNDNTEFSSTK-----YKANSLEDEDESLAKE 922
QY 644 AHR-EVIN-----SSTEGLLNIDKIRKILSGYIIEIETEGLK 682
Db 923 LTKAEVYDKLDEESEDSEYEQDYADPEPGNDEGSNENIYVGTGK-----DTLGIV 972
QY 683 EYVNDRYDMLNISLQDQKTFIDFKKYNDKLLPLYSNPYKVVYAVYKENTLIN 738
Db 973 EPENEKVN-----KVHEETLFEANVSSVSVQVONKMDMTVIN 1010

RESULT 15
T18351
Impl protein - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 07-Dec-1999
C:Accession: T18351
R:Jensen, L.T.; Ladefoged, S.; Birkelund, S.; Christensen, G.
A:Title: Selection of Mycoplasma hominis Pcz1 deletion mutants by cultivation in the pre
A:Reference number: Z18884; MUID:95369882
A:Accession: T18351
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1051 <JEN>
A:Cross-References: EMBL:U21961; NID:g790241; PID:g790242; PIDN:AAA81012.1
C:Genetics:
A:Gene: Impl
A:Genetic code: SGC3

Query Match 4.6%; Score 180; Db 2; Length 1051;
Best Local Similarity 19.7%; Pred. No. 0.14;
Matches 170; Conservative 137; Mismatches 324; Indels 232; Gaps 43;
QY 2 KRRVILPLAAL-----STILVSTSGNLEVIQAEVKQKNRLNSESSESSQ--GLIGY- 51
Db 4 KKKKIALPLAILCGGLAIATTAIARRRHAKIANOSQKENVLQNFKKLOKQNLNGLGK 63
QY 52 YFSDLN-FQAPVYVTSSTGDLSPS--SELENIPSENOYQSAIWSGFIKRV----- 101
Db 64 IYVEINVFHQEVL-----QGSLLKNNKSETKALIEETLRKDAITTLISKINGINKE 119
QY 102 -----KSDY---TATSADNHVYMWDDQEVYINKASNKIRLEKGLYQIK-- 146
Db 120 EFAKFNEIKKQLEYIKNEISKOYEYEH-----KQNIENELNXYPTISLE-STLIEQNA 173
QY 147 -----IOYQENPTEKGLDFKLYWTDSONKKE---VSSDNLOLPELKQSSNRKKRST 198
Db 174 TNNILKILNISTEKDN-----IDNLNAKEQLKASISQANOLLQDSDNDSEIARAKK- 226
QY 199 SAGPTVPDRDNDGIPDSLEVE-----GYVDVKNKRTFLSPWISNIHEKGLTKY 248
Db 227 -----SLDAEIKKANQAVASNNMTASQSAKSSLDAAKVAETTK--LET 268
QY 249 KSPPEKWTASDPYSDPEKVTGRIDKNVSPEARHP---LVAATPIVHVMENIILSKNE 304
Db 269 NKDKEA-----KFNEKQORNDQOEFINTKNPNYSELISQITSKR-DKSNVYDSSN 321
QY 305 DQSTQNTDSEFRTISKNTSRTHTSEVHGNAEVAHANTSTSRHTSEVHGNAEVAID 364
Db 322 KSDIESANTELKQALAKANDKVOADNLAKSIKQNLNNSVSNANT-----LSALTD 373
QY 365 HSLSLAGERTWAE-----TGLNTADTARLNMIRVYNTGTATVYVLPPTSILVGLKNQ 418
Db 374 KDMTIOAKKELEKEIQKANOAKISNNTASMOQSAKSSLDAAKVAETTKLETFN--KQKEA 431
QY 419 TLATIKAKENQLOIILAPNNYTP-----SKNLAPIALNAQDDFSSPTIM 464
Db 432 KFNEKQORNDQOEFINTKNPNYSELISQITSKRDSKSNVYDSSNKSQIESA--NTE 488
QY 465 YNOFLEKTKQRLDLD--QVYGNIAVYNEENGHRVVDGSMSEVL-PQIOETTARIIIF 522
Db 489 LKQALNTAKAKSSIDNELPLKNDQSKIEEGEPTR-NTNFSVT----- 532
QY 523 NGKDLNLYERRIAAVNPSPLETTK---PDMTLKALKIAGFNEPNNGN----- 568
Db 533 -----SSKLETTKKNLAEIETRKADAIK-----NPNSSKQALKOSSQ 570
QY 569 LOYQ-----KDTEPDEFNPQOQTSQNIKNQDLAELNATNIV--TVLDKIK-----LN 613
Db 571 VQKIGNELLTITE-EPGKVEYTKNSNIGYRLFKLAQEQQRNNSVDPLKNAWEKQTLIS 629
QY 614 AKNNIILRDKRFYDRNNIIVAGDESVK-----BAH-REVINSSTEGLLNIDKIRK 666
Db 630 KQKRLQNSTKDYLTQSTMSSTOESTIKKYIVNIQAHINNLNSOYR---LEADKLIAN 686
QY 667 ILSGY-IYVEIETEGLEKVIYNDRYDMLNI-SSLRQDGKTFIDFKKYNDKRLPL---YIS 719
Db 687 MKRGYQKVGIESLQKKQDMLMD--SVLSVDDSLAKD-----FNKALRYLYGVGYTK 735
QY 720 NPKYKVVYAVYKENTLINPSEN 742
Db 736 NP--PVSSWPFINKRNRISIENYQN 756

Search completed: December 2, 2001, 13:51:18
Job time: 313 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:48:22 : Search time 112.1 Seconds
(Without alignments)
504,835 Million cell updates/sec

Title: US-09-747-521-4
Perfect score: 3913
Sequence: 1 MKRRKVLPLMALSTILVSS.....TSTNGIKKILFSSKKGYEIG 764

Scoring table: BIOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_1101.*
1: /SID8/gcgdata/geneseq/genesqp/AA1980.DAT.*
2: /SID8/gcgdata/geneseq/genesqp/AA1981.DAT.*
3: /SID8/gcgdata/geneseq/genesqp/AA1982.DAT.*
4: /SID8/gcgdata/geneseq/genesqp/AA1983.DAT.*
5: /SID8/gcgdata/geneseq/genesqp/AA1984.DAT.*
6: /SID8/gcgdata/geneseq/genesqp/AA1985.DAT.*
7: /SID8/gcgdata/geneseq/genesqp/AA1986.DAT.*
8: /SID8/gcgdata/geneseq/genesqp/AA1987.DAT.*
9: /SID8/gcgdata/geneseq/genesqp/AA1988.DAT.*
10: /SID8/gcgdata/geneseq/genesqp/AA1989.DAT.*
11: /SID8/gcgdata/geneseq/genesqp/AA1990.DAT.*
12: /SID8/gcgdata/geneseq/genesqp/AA1991.DAT.*
13: /SID8/gcgdata/geneseq/genesqp/AA1992.DAT.*
14: /SID8/gcgdata/geneseq/genesqp/AA1993.DAT.*
15: /SID8/gcgdata/geneseq/genesqp/AA1994.DAT.*
16: /SID8/gcgdata/geneseq/genesqp/AA1995.DAT.*
17: /SID8/gcgdata/geneseq/genesqp/AA1996.DAT.*
18: /SID8/gcgdata/geneseq/genesqp/AA1997.DAT.*
19: /SID8/gcgdata/geneseq/genesqp/AA1998.DAT.*
20: /SID8/gcgdata/geneseq/genesqp/AA1999.DAT.*
21: /SID8/gcgdata/geneseq/genesqp/AA2000.DAT.*
22: /SID8/gcgdata/geneseq/genesqp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3800	97.1	764	21	AAV56958 B. anthracis prote
2	3800	97.1	764	22	AAAB47306 Wild type B. anthr
3	3699.5	94.5	763	21	AAV56960 B. anthracis TPA-P
4	3667	93.7	735	15	AAAB60179 Protective antigen
5	3667	93.7	736	21	AAV56959 B. anthracis MAT-P
6	3622.5	92.6	903	15	AAAB60183 PA(1-725)----Huma
7	3490	89.2	719	15	AAAB60193 Modified protectiv
8	2812	71.9	569	21	AAV56961 B. anthracis PA63
9	790.5	20.2	884	17	AAAB91239 B. cereus VIP1A(a)
10	789.5	20.2	884	18	AAAB19509 B. cereus VIP1A(a)
11	789.5	20.2	884	19	AAAB67112 100 kDa VIP1A(a) t

12	789.5	20.2	1346	17	AAAB91245 VIP2A(a) and VIP1A
13	789.5	20.2	1346	18	B. cereus VIP1A(a)
14	789.5	20.2	1346	18	AAAB67233 VIP1A(a)/VIP2A(a)
15	787.5	20.1	880	21	AAAB60224 Bacillus thuringie
16	787	20.1	881	21	AAV59227 MIS toxin from B.
17	784.5	20.1	884	15	AAAB63793 Bacillus cereus 10
18	777	19.9	1338	17	AAAB91247 VIP2A(a)-VIP1A(a)
19	777	19.9	1338	18	AAAB19520 Maize optimised-B.
20	777	19.9	1338	19	AAAB67311 VIP2A(a)/VIP1A(a)
21	776.5	19.8	784	21	AAV82948 MIS toxin of Bacil
22	775	19.8	852	17	AAAB91246 VIP1A(a) protein w
23	775	19.8	852	18	AAAB19516 Maize optimised-B.
24	775	19.8	852	17	AAAB67233 Maize optimised VI
25	772	19.7	860	21	AAV59282 MIS-8 toxin from B
26	766	19.6	834	17	AAAB91242 B. thuringiensis V
27	766	19.6	834	18	AAAB19512 B. thuringiensis V
28	766	19.6	834	19	AAAB6722 Vegetative insect
29	596.5	15.2	425	19	AAAB60222 Bacillus thuringie
30	587	15.0	667	17	AAAB91240 B. cereus VIP1 pro
31	586	15.0	667	18	AAAB19510 B. cereus 80 kd VI
32	586	15.0	667	19	AAAB6713 80 kDa VIP1A(a) to
33	584	14.9	667	15	AAAB63794 Bacillus cereus 80
34	541	13.8	357	19	AAAB60228 Bacillus thuringie
35	538	13.7	357	19	AAAB60231 Bacillus thuringie
36	521	13.3	357	19	AAAB60218 Bacillus thuringie
37	521	13.3	357	19	AAAB60226 Bacillus thuringie
38	503	12.9	357	19	AAAB60227 Bacillus thuringie
39	450	11.5	327	19	AAAB60220 Bacillus thuringie
40	447	11.4	327	19	AAAB60223 Bacillus thuringie
41	427.5	10.9	348	19	AAAB60219 Bacillus thuringie
42	425.5	10.9	348	19	AAAB60229 Bacillus thuringie
43	419.5	10.7	347	21	AAAB60221 Bacillus thuringie
44	419.5	10.7	347	21	AAV59276 Toxin from B. thur
45	396	10.1	333	19	AAAB60230 Bacillus thuringie

ALIGNMENTS

RESULT 1
ID AAV56958 standard; protein: 764 AA.
AC AAV56958:
XX
XX 25-APR-2000 (first entry)
XX
XX B. anthracis protective antigen (PA) protein.
DE
XX Bacillus anthracis protein: protective antigen; PA; MAT-PA; TPA-PA;
KW tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.
XX
XX Bacillus anthracis.
OS
XX
XX WO200002522-A2.
PN
XX
XX 20-JAN-2000.
PD
XX
XX 09-JUL-1999; 99WO-US15568.
PE
XX
XX 10-JUL-1998; 98US-0092416.
PR
XX
XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
PA
XX Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;
PI
XX WPI; 2000-182165/16.
XX N-PSDB; AAZ56874.
DR
XX
XX Recombinant DNA construct useful as vaccines for anthrax, in producing
PT host cells for analyzing the drugs and agents inhibiting anthrax
PT
XX
XX Disclosure; Page 33; 35pp; English.
PS


```
QY 61 PAVTSTTGDLSPSELENIPSENOYFOSAIMSGFIKKKSEDEYFATSADNHVTMW 120
XX |||
PI 61 pmvtsatctgdslpselenipseengyfgsalwsgfikkksdeyfatcsadnhvtmw 120
XX |||
QY 121 DDOEVINKASNNKIRLEKRLYQIKIYORENPTKEGDLFKLYWDSQNKKEVISSDNL 180
XX |||
DB 121 ddegevinakasnnkirkrlqyikqyrenptekyldfklywdsqnkkevi ssdnl 180
XX |||
QY 181 QPELQKSSNSKKRSTSGPTVPDRNDGIPDSLEVEGYTDVKKRFFLSWISNIN 240
DB 181 qpelqkssnsrkrstsgptvpdrndgipdsleveytdvknkrfflspwlsninh 240
XX |||
QY 241 EKKGLFKYKSSPEKMWSTASPDYSDFEKVTGRIRKNSPEARHPVLVAAPYIVHDMENITL 300
DB 241 ekkglfkyksspekswstasdpdydfekvgrirknvspearhplvaaypivhdmennll 300
XX |||
QY 301 SKNEDOSTONTDESETRTISKNTSTSRTHSEVHGNAEVHANTSTSRTHSEVHGNAEVHA 360
DB 301 sknedgstontdsetrtiskntstsrthsevhgnaevhasffdigsvsaagfsnnsst 360
XX |||
QY 361 VALDHSLSLAGERTVAETMGLNADTARLANIRYVNTGAPLYNVLPPTSLVGNQTL 420
DB 361 valdhsllsagertvaetmglncadtarlanirvntgapiynvlptslvlgknqtl 420
XX |||
QY 421 ARTIKAKENOLSOILAPNNYPSKNLAPIALNADDFSSPTITMNYNOFLELEKTKOLRLD 480
DB 421 artikakenqslilapnnypsknlapia lnaddfsscpitmmnyftelektkqlrld 480
XX |||
QY 481 TDQVYGNATYFENGSRVAVDGSNMSEVLPOIETTARIIFNGKDLNVERRIAIVNPS 540
DB 481 tdqvygnatynfengsravravdgsnmsevlpoiettariifngkdlnverriaevnps 540
XX |||
QY 541 DPLETRKPMTLKEALKIAGFNEPNGNLOYGKDTIEFPNPDQOTSONIKQALNALNA 600
DB 541 dpletckpmtlkealkiafgnepngnllygqkdtiefnfdqgtsqnlknqlaelna 600
XX |||
QY 601 TNYTVLDRKIKLNAKNNILIRDRKFHYDRNNIIVAGDESVYKKAHREVINSSTEGLLNIT 660
DB 601 tnytvldrkiklnaknnilirdkrfhydrnniivagadesvvykkahevinssteglllni 660
XX |||
QY 661 DKDIRKILSGYIEIEDTEGLKEVINDRYDMLNITSLRODGKTFIDKKYNDKLLPLYISN 720
DB 661 dkdirkilsgyieiedteglkevindyrdmlnitslrodgktfidckkynndkplyisn 720
XX |||
QY 721 PNKVVAVYATKENTIIINPSENGDTSNGIKKILIFSKKGEYEG 764
DB 721 pnkvvavyatkentiiinpsengdstngikilifskkyeg 764
XX |||
RESULT 3
AA56960
ID AAY56960 standard; protein; 763 AA.
AC AAY56960;
XX
DT 25-APR-2000 (first entry)
XX
DE B. anthracis TPA-PA protein.
XX
KM Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA;
XX tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.
OS Bacillus anthracis.
PN WO200002522-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US15568.
XX
PR 10-JUL-1998; 98US-0092416.
XX
```

```
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL.
XX
DR WPI; 2000-182165/16.
XX
PT Recombinant DNA construct useful as vaccines for anthrax, in producing
XX host cells for analyzing the drugs and agents inhibiting anthrax
PS Disclosure; Page 32; 35pp; English.
XX
CC The invention provides a recombinant DNA construct that comprises a
CC vector and at least one nucleic acid (or its fragment) encoding a
CC combination of Bacillus anthracis proteins, selected from protective
CC antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA
CC (PA with its secretory signals replaced with those of tissue plasminogen
CC activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine
CC for anthrax and in producing infectious alpha virus particles. These
CC particles, expressing the B. anthracis proteins are useful also as
CC vaccines for anthrax. Host cells transformed with the construct are
CC useful for analyzing the effectiveness of drugs and agents that inhibit
CC anthrax or B. anthracis proteins. The present sequence represents a
CC B. anthracis TPA-PA protein.
XX
SQ Sequence 763 AA:
XX
Query Match 94.5%; Score 3699.5; DB 21; Length 763;
Best Local Similarity 95.4%; Pred. No. 7.5e-231;
Matches 727; Conservative 7; Mismatches 25; Indels 3; Gaps 1;
QY 3 KRKVLIRLMLSTLIVSTGMLVYIOAEVKNLNLNSESOGGLGYRSDNLFCQPM 62
DB 5 krglcvlllvcgavfvaas---eviaevkqenllnesessqgllyyfsdlnlfcqpm 61
XX |||
QY 63 VVTSSTGDLSPSELENIPSENOYFOSAIMSGFIKKKSEDEYFATSADNHVTMWDD 122
DB 62 vvtsstgdlspseelenipseengyfgsalwsgfikkksdeyfatcsadnhvtmwdd 122
XX |||
QY 123 QEVINKASNSNKIRLEKRLYQIKIYORENPTKEGDLFKLYWDSQNKKEVISSDNLQ 182
DB 122 qevinkasnnkirkrlqyikqyrenptekyldfklywdsqnkkevi ssdnlq 181
XX |||
QY 183 PELKQSSNSKKRSTSGPTVPDRNDGIPDSLEVEGYTDVKKRFFLSWISNINHEK 242
DB 182 pelqkssnsrkrstsgptvpdrndgipdsleveytdvknkrfflspwlsninh 241
XX |||
QY 243 KGLTFKYSPEKMWSTASPDYSDFEKVTGRIDKNVSPARHPVLVAAPYIVHDMENITLSK 302
DB 242 kgltfkysspekswstasdpdydfekvgridknvnspearhplvaaypivhdmennllsk 301
XX |||
QY 303 NEDOSTONTDESETRTISKNTSTSRTHSEVHGNAEVHANTSTSRTHSEVHGNAEVHAVA 362
DB 302 nedgstontdsetrtiskntstsrthsevhgnaevhasffdigsvsaagfsnnssteva 361
XX |||
QY 363 IDHSLSLAGERTVAETMGLNADTARLANIRYVNTGAPLYNVLPPTSLVGNQTLAT 422
DB 362 idhsllsagertvaetmglncadtarlanirvntgapiynvlptslvlgknqtlat 421
XX |||
QY 423 IKAKENOLSOILAPNNYPSKNLAPIALNADDFSSPTITMNYNOFLELEKTKOLRLDPTD 482
DB 422 ikakenqslilapnnypsknlapia lnaddfsscpitmmnyftelektkqlrldptd 481
XX |||
QY 483 QVYGNATYFENGSRVAVDGSNMSEVLPOIETTARIIFNGKDLNVERRIAIVNPSDP 542
DB 482 qvygnatynfengsravravdgsnmsevlpoiettariifngkdlnverriaevnpsdp 541
XX |||
QY 543 LETTRKPMTLKEALKIAGFNEPNGNLOYGKDTIEFPNPDQOTSONIKQALNELANTN 602
DB 542 letckpmtlkealkiafgnepngnllygqkdtiefnfdqgtsqnlknqlaelnntn 601
XX |||
QY 603 IYTVLDRKIKLNAKNNILIRDRKFHYDRNNIIVAGDESVYKKAHREVINSSTEGLLNIDK 662
XX |||
```

|||||
Db 602 lvtvdtklknkmlllrdrfhydrnlaavadesvvrkaehrevlnssteglllnldk 661
OY 663 DIRKLSGYIEIEDTEGLAKEVINDRYDMLNSSLRODGKTFIDPKKYNDRKLPYISMPN 722
Db 662 dirklsgylveledteglakevindrmdlnslsrldgkctfidtkkynndkplyismpn 721
OY 723 YKVNVAATKENTTIINPSENGDSTNGIKKILIFSKGYEIG 764
Db 722 ykvnvavtkencllnpsengdstnglkkllfsskgyeig 763

RESULT 4

AAR60179 standard; Protein: 735 AA.

AAR60179;

03-APR-1995 (first entry)

Protective antigen of *Bacillus anthracis*.

Anthrax; *Bacillus anthracis*; fusion protein; targeting; targeting; pathogen;

Intracellular; HIV; human immunodeficiency virus; toxin.

Bacillus anthracis.

W09418332-A.

18-AUG-1994.

14-FEB-1994; 94WO-US01624.

12-FEB-1993; 93US-0021601.

25-JUN-1993; 93US-0082849.

(USSH) US DEPT HEALTH & HUMAN SERVICES.

Atora N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;

WPI: 1994-279753/34.

N-PSDB: AAQ70180.

Nucleic acid encoding anthrax toxin fusion protein - useful for targeting toxin to specific cells, eg for killing tumour cells or HIV-infected cells

Disclosure: Page 81-83; 124pp; English.

The sequence encoding the protective antigen of *Bacillus anthracis* may be used in the construction of a nucleic acid which encodes a fusion protein comprising the anthrax protective antigen binding domain of the native anthrax lethal factor and a sequence encoding an actively inducing domain of a second protein. The fusion proteins are useful for the specific killing of tumour cells or the killing of cells infected with intracellular pathogens, especially HIV.

Sequence 735 AA;

Query Match 93.7%; Score 3667; DB 15; Length 735;

Best Local Similarity 97.6%; Pred. No. 8.9e-229;

Matches 717; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

OY 30 EYKQENRLINSESSSOGILGYFSDLNFOAPMVYTSSTGDLSTPSSLENIPEENQYF 89
Db 1 evkgenrllneesssgqllgyfisdlnqapmvvtstcgslspsselenipengyf 60
OY 90 QSAISGEIKVKKSDYETATADNHVTMWVDDQEVINKASNSNKRIRLEKGRLOYIKIY 149
Db 61 qsaltsgefkvkkseytatsadnhvtmwvddgevinkasnsnkrlrlekgrloyiklqy 120

OY 150 QRENTEKGLDFKLYWTDSONKKEYISSDNQLPELKOKSSNSRKRKSTAGPYVPPDRN 209
Db 121 qrenptekglidfklywtdsqnkkeyissdnqlpelkksnsrkrkstasagpyvpddrn 180
OY 210 DGIPDSLEVEGYTDVKKRPFELSPWISNHEKKGLTFYKSSPEKWSASDPYDPEKVT 269
Db 181 dgipdslevegtydvkkrpfelspwlnhekkgltyksspekwsasdpysdfekev 240
OY 270 GRIDKNVSPPEARHPLVAAPYIVHVMENIILSKNEDOSTQNTDSETRTISKNTSRTHT 329
Db 241 gridknvspearhplvaapylvhwmdeniilsknedgstqntdsertiskntsrtcht 300
OY 330 SEVHGNAEVHANTSTSRHTSEVHGNAEVHAAVADHSLSLGERMATKMLNADTARL 389
Db 301 sevhgnaevhasfidiagvsagfnsnsstvaldhsislagertwaelmglnadetarl 360
OY 390 NANTRYVNTGTAPYINVLPTSLVGRKQOTLATIRAKENOLSLAPNNYPSKNLAPIA 449
Db 361 nantryntgcaplynvlptslvgknqlatlrakenglsqllapnnypsknlapia 420
OY 450 LNAODPFSSPTITMNYNOFLELEKTKQLRLDQVYGNATVYNENGRVRVDTGSMNSEV 509
Db 421 lnaoddfsstptitlmnyoflelektkqlrldtqygniatynfengrvrvdtgsnsew 480
OY 510 LPQIOETARILIFNGKDLNLYERRIAAVNPSDPLETTPDMTLKALKIAGFNPENGNL 569
Db 481 lpqioetarilifngkdlnlverriaavnsdpdlettkpdmtlkaalkiagfnepgnl 540
OY 570 QYOGKDIPEFPNFPOQTSQINIKNOLAEINATNYTVVDKIKLANKNILIRDRKFRHFR 629
Db 541 qyogkdipefndfngqtsqiniknglaelnatnytvvdiklankmlllrdrkfrhfr 600
OY 630 NNIAVGADESVYKKAHREVINSSTEGLLNLIDKDIRKILSGYIEIEDTEGLKEYINDRY 689
Db 601 nniavgadesvvrkaehrevlnssteglllnldkdrkllsgylveledteglakevindr 660
OY 690 DMLNSSLRODGKTFIDPKKYNDRKLPYISMPNPNVYVYATKENTTIINPSENGDSTNG 749
Db 661 dmlnsslrdgkctfidtkkynndkplyismpnykvavtkencllnpsengdstng 720
OY 750 IKKILIFSKKGYEIG 764
Db 721 ikkllfsskgyeig 735

RESULT 5

AAY56959 standard; protein: 736 AA.

AAY56959;

25-APR-2000 (first entry)

B. anthracis MAT-PA protein.

Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA; tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.

Bacillus anthracis.

W0200002522-A2.

20-JAN-2000.

09-JUL-1999; 99WO-US15568.

10-JUL-1998; 98US-0092416.

(USME-) US MEDICAL RES INST INFECTIOUS DISEASES.

Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;

WPI: 2000-182165/16.

DR N-PSDB; AA556875.

XX Recombinant DNA construct useful as vaccines for anthrax, in producing
PT host cells for analyzing the drugs and agents inhibiting anthrax
XX
PS Disclosure; Page 34; 35pp; English.

XX The invention provides a recombinant DNA construct that comprises a
CC vector and at least one nucleic acid (or its fragment) encoding a
CC combination of Bacillus anthracis proteins, selected from protective
CC antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA
CC (PA with its secretory signals replaced with those of tissue plasminogen
CC activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine
CC for anthrax and in producing infectious alpha virus particles. These
CC particles, expressing the B. anthracis proteins are useful also as
CC vaccines for anthrax. Host cells transformed with the construct are
CC useful for analyzing the effectiveness of drugs and agents that inhibit
CC anthrax or B. anthracis proteins. The present sequence represents a
CC B. anthracis MAT-PA protein.

XX Sequence 736 AA:

Query Match 93.7%; Score 3667; DB 21; Length 736;
Best Local Similarity 97.6%; Pred. No. 9e-229;
Matches 717; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

```

OY 30 EVKQENRLNSESSESSGGLGYFSDINFOAPMVVTSSTGDLSPSELENIPSENOYF 89
DB 2 evkqenrlinesesssgllgyfisdlnfqapmvvtstgdlspseleinpsengyf 61
OY 90 QSAIWSGFIKVKKSDDEYTFATSADNHVTMWVDDQEVINKASNSKIRLEKGRLYQIKIY 149
DB 62 qsaIwsgfIkVksdEYtFatsAdNHvTMwVDDQEvInKASnSkIRlEkGrLyqIkIy 121
OY 150 QRENPTKGLDFKLYWTDSONKKEVVISDNIQLPELKOKSSNSKRRKSTAGPVPPDRN 209
DB 122 grenptekglDfKlywTdSoNkKevViSSdNIqLpELkOKSSnSKRRkSTAGpVPpDRn 181
OY 210 DGIPDSELENGYIVDVAKNKTFTLSPWISNTEHEKGLTKYSSPEKWTASDPISDEPKVT 269
DB 182 dgIpdselEngyIvDvAKnKtFTlSpWISnTEhEkGLtKYSSpEkWtASdPiSdEPkVt 241
OY 270 GRIDKNVSPFARHPLVAAYPIVHVDMEITLSKNEDOSTONTSEYTIKKNSTSTRTHT 329
DB 242 gridknvspFARhPlVAayPIvHvDMEITlSKNEdOSTONTSEyTIKKnSTSTRTht 301
OY 330 SEVHGNAEVHANTSTSTRTHTSEVHGNAEVHAVAIDHSLSLAGERTMAETGLNTADTARL 389
DB 302 sevhgnaevhAntstSTRThtSEvHGNAEvHAVAIDHsLSLAGERTMAETGLNTADtARl 361
OY 390 NMIRRVNVTARIVNVLPTTSVLGKNQTLATIKAKENLSQILAPANNYPPSKNLAPIA 449
DB 362 nmiRrvNvTARiVnVLpTTSvLGKnQTLATIKAKENLSQILAPANNyPPSKnLAPIa 421
OY 450 LNAODESSPTPIRMANTNOPIELEKTKOLRLDQOVYGNITVNEGRVAVDGSNMSEV 509
DB 422 lnaodessptPIrMANtNOPIELEkTKOLrLDQOVyGNITvNEGRvAVDGSNMSEv 481
OY 510 LPOIETTARIIFNGDLNLVERRIAANVPSDPLETTKPMTLKEAKIAFGFNEPNCNL 569
DB 482 lpoIeTtARiIFngDLnLVERrIAANvPSDPLEtTKPMtLKEAKIAFGfNEpNCnl 541
OY 570 QYQGRKITEFDNFDOOTSINIKNOALNATNITYVLDIKILAKANNIIRDKRFHYDR 629
DB 542 qyQGRkITEfdNFdOOTSINiKNOALNATNITyVLDIKILAKANNiIRdKRFHydr 601
OY 630 NNIAVGADESVEKAREVINSSTEGLLNIDDIRKILSGYIEIDTGLKRVINDRY 689
DB 602 nnIaVGADESvEkAREvINSstEGllNIddIRKiLSgYIEIdTGLKRVINdRy 661
OY 690 DMLNSSLRODGKFTIDFKKYNKDLPLYISNPNKYVAVTKENTIINSENGDSTNG 749
DB 662 dmlNssLRoDGkFTIDfKKYnKDLpLYISnPNKYvAVtKEntIINSEngDStNG 721

```

DB 662 dmlnsslrqdgkftidfkkyndkplyisnpnykvnvavtkeintiinsengdstng 721
OY 750 IKRLIFSKKGYENG 764
DB 722 ikrlifskkyeng 736

RESULT 6

AA60183
ID AAR60183 standard; Protein: 903 AA.

AC AAR60183;
XX

DT 04-APR-1995 (first entry)
XX

DE PA(1-725)-----Human CD4 fusion protein coding sequence.
XX

KW Anthrax; Bacillus anthracis; fusion protein; lethal factor;
KM protective antigen; cell killing; targeting; targeting; pathogen;
KW Intracellular; HIV; human immunodeficiency virus; toxin;
KM Pseudomonas; exotoxin.
XX

OS Bacillus anthracis.
OS Homo sapiens.
XX

PN WO9418332-A.
XX

PD 18-AUG-1994.
XX

PF 14-FEB-1994; 94WO-US01624.
XX

PR 12-FEB-1993; 93US-0021601.
PR 25-JUN-1993; 93US-0082849.
XX

PA (US93) US DEPT HEALTH & HUMAN SERVICES.
XX

PI Atora N, Kimpel K, Leppia SH, Nichols PJ, Singh Y;
XX

DR WPI: 1994-279753/34.
XX

DR N-PSDB; AA070184.
XX

PT Nucleic acid encoding anthrax toxin fusion protein - useful for
XX targeting toxin to specific cells, eg for killing tumour cells
XX or HIV-infected cells

PT Nucleic acid encoding anthrax toxin fusion protein - useful for
XX targeting toxin to specific cells, eg for killing tumour cells
XX or HIV-infected cells

PS Disclosure; Page 100-103; 124pp; English.
XX

XX This sequence is a fusion protein comprising amino acid residues
CC 1-725 of the anthrax protective antigen protein and residues 1-178
CC of human CD4, the portion which binds to gp120 on HIV infected
CC cells. Such fusion proteins may be useful for the specific killing
CC of tumour cells or the killing of cells infected with intracellular
CC pathogens, especially HIV, depending on their components.
XX

XX Sequence 903 AA:
XX

Query Match 92.6%; Score 3622.5; DB 15; Length 903;
Best Local Similarity 96.7%; Pred. No. 9.1e-226;
Matches 710; Conservative 3; Mismatches 18; Indels 3; Gaps 1;

```

OY 30 EVKQENRLNSESSESSGGLGYFSDINFOAPMVVTSSTGDLSPSELENIPSENOYF 89
DB 1 evkqenrlinesesssgllgyfisdlnfqapmvvtstgdlspseleinpsengyf 60
OY 90 QSAIWSGFIKVKKSDDEYTFATSADNHVTMWVDDQEVINKASNSKIRLEKGRLYQIKIY 149
DB 61 qsaIwsgfIkVksdEYtFatsAdNHvTMwVDDQEvInKASnSkIRlEkGrLyqIkIy 120
OY 150 QRENPTKGLDFKLYWTDSONKKEVVISDNIQLPELKOKSSNSKRRKSTAGPVPPDRN 209
DB 121 grenptekglDfKlywTdSoNkKevViSSdNIqLpELkOKSSnSKRRkSTAGpVPpDRn 180

```

```

QY 210 DGIPDSLEVEGYTVVKNKRTFLSPWISNIHEKKGLTKYKSSPEKWSASDPYSPFEKVT 269
DB 181 dgiplslevegyltvvknkrtflspwlsnlnhekgltkyspekwsasdpysdfekev 240
QY 270 GRIDNVSPEARHPLVAAYPIVHVDMENILISKNEOSTOMTDETRTISKNTSRTHT 329
DB 241 gridnvspearhplvaayplivhvdmenilisknedgstqntdsetrtlskntscrtht 300
QY 330 SEVHGNAEVHANTSTSRHTSEVHGNAEVHAYADHSLSLAGEERTWAGTGLNTADTARL 389
DB 301 sevhgnaevhasffdiggsvasgafsnstsvaidslslageretwagtmjntadtarl 360
QY 330 NANIRYVNTGAPRYNVLPPTSLVYGNKQTLATIKAKENQSLIAPNNYPSKMLAPIA 449
DB 361 nanirylvntgaprynvlpptslvlgknqtlatlakakenqslgapnnypskmlapia 420
QY 450 LNAODFESSPTITMYNOFLFLEKTKOLRLDPOYGNATYNEFNGRVRVDTGSNMSRV 509
DB 421 lnaodfessptitmyngflflekckqlrltdqvygnatlynefngvrvtqgsnmsrv 480
QY 510 LPQIDETARIIFNGKDLNVERRIAAYNPSPDLETTKPDMTLKEALKIAFGFNPNGL 569
DB 481 lpgidettarilifngkdlnverriaavnpdpletctkpdmtlkealkiafgfnepngl 540
QY 570 QYQGDITREFPDPNPOQTSQNIKNQALNATNTYTVDKIKLNKKNMILIPDKRFHIDR 629
DB 541 qyqgdtrefdpnpoqtsqnknqlaelnactnlytvldkiklnaknmililrdrkrfhydr 600
QY 630 NNIAVGADEVYKREAHREYVINSSTEGELNLNDIKDRILISGYIVEDETEGLEKVINDRY 689
DB 601 nniavgadesvvykrehrevinssteglllnldkdrilksyiveetedeglekeyindry 660
QY 690 DMLNLSLRQDGKTFIDFKKYNDKLPXYISNPYKVNVAVTKEMTIINPSENGDTSNG 749
DB 661 dmlnlsrlrqdgktfidfkkyndkrlpxyisnpnykvnyavtkentlinpsengdstng 720
QY 750 IKKIL---IFSKKG 760
DB 721 ikkilkvvlqkg 734

```

RESULT 7
AAR60193 standard; Protein: 719 AA.

AC AAR60193:

DT 04-APR-1995 (first entry)

DE Modified protective antigen of Bacillus anthracis.

KW Anthrax; Bacillus anthracis; fusion protein; lethal factor;
protective antigen; cell killing; targeting; targeting; pathogen;
intracellular; HIV; human immunodeficiency virus; toxin.

OS Bacillus anthracis.

XX WO9418332-A.

XX 18-AUG-1994.

PF 14-FEB-1994; 94WO-US01624.

XX 12-FEB-1993; 93US-0021601.

PR 25-JUN-1993; 93US-0082849.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Atora N, Klimpel R, Lepia SH, Nichols PJ, Singh Y;

XX WPI: 1994-279753/34.

DR N-PSDB: AAQ70189.

XX

PT Nucleic acid encoding anthrax toxin fusion protein - useful for
PR targeting toxin to specific cells, eg for killing tumour cells
or HIV-infected cells

PS Example 6; Page 114-115; 124pp; English.

CC The sequence encoding the protective antigen of Bacillus anthracis
CC may be used in the construction of a nucleic acid which encodes a
CC fusion protein comprising the anthrax protective antigen binding
CC domain of the native anthrax lethal factor and a sequence encoding an
CC activity inducing domain of a second protein. Such fusion proteins
CC are useful for the specific killing of tumour cells or the killing of
CC cells infected with intracellular pathogens, especially HIV,
CC depending on the second component. The protective antigen and other
CC toxins require proteolytic cleavage to acquire activity. Since some
CC cells infected with an intracellular pathogen possess an active
CC protease with quite a narrow substrate specificity e.g. HIV, the
CC protease cleavage site found in the native toxin is replaced with an
CC intracellular pathogen specific protease site (See AAR60184-88). The
CC protease in cells that are infected with an intracellular pathogen
CC cleaves the modified toxin which is then rendered active and kills
CC the cell. This sequence is a modified Bacillus anthracis protective
CC antigen which has the amino acids originally at positions 162-171
CC replaced with the HIV protease cleavable sequence described in
CC AAR60186.

SO Sequence 719 AA;

Query Match 89.2%; Score 3490; DB 15; Length 719;

Best Local Similarity 93.1%; Pred. No. 2.4e-217; Mismatches 21; Gaps 2;

Matches 688; Conservative 6; Mismatches 21; Indels 24; Gaps 2;

```

QY 30 EYKQENRLNSESSESSQGLGYFSDLNFOAPMVYTSSTGDLSPSELENIPESENGYF 89
DB 1 evkqenrlneesssglgyfisdlnfqpemvtsstgdlspselelpeengyf 60
QY 90 QSAIWSGFIRYKKSDEYTPATSDADNHVTWVDDQEVINKASNSNKRIRLEKGLYQIKQY 149
DB 61 qsaIwsGfIrYkKsDeYtPaTsAdNhVtWvDdQeViNkAsNsNkRiRlEkGlYqIkQy 120
QY 150 QRENTEKGLDPFKLWTSQNKKEVYISSNDOLPELKKOKSSNS----KKRSTSGRPVP 205
DB 121 qrenTEkglDpFkLwTsQnKkEvYiSSNdOlPeLkKsShtatImngRgVflIqgRPVP 180
QY 206 DRDNDGIPDSLEVEGYTVVKNKRTFLSPWISNIHEKKGLTKYKSSPEKWSASDPYSPDE 265
DB 161 drdndgipdslevegyltvvknkrtflspwlsnlnhekgltkyspekwsasdpysdf 240
QY 266 EKVYGRIDKNVSPEARHPLVAAYPIVHVDMENILISKNEOSTOMTDETRTISKNTS 325
DB 241 ekvtgrIdknvSearHplVaaYpIvHvDMeNiLIsKNeOsToMtDeTrtIsKntS 300
QY 326 RHTSEVHGNAEVHANTSTSRHTSEVHGNAEVHAYADHSLSLAGEERTWAGTGLNTAD 385
DB 301 rhtseVhgnaevHasffdiggsvasgafsnstsvaidslslageretwagtmjntad 360
QY 386 TARLANIRYVNTGAPRYNVLPPTSLVYGNKQTLATIKAKENQSLIAPNNYPSKML 445
DB 361 tarlanirylvntgaprynvlpptslvlgknqtlatlakakenqslgapnnypskml 420
QY 446 AFIALNAODFESSPTITMYNOFLFLEKTKOLRLDPOYGNATYNEFNGRVRVDTGSN 505
DB 421 afialnaodfessptitmyngflflekckqlrltdqvygnatlynefngvrvtqgsn 460
QY 506 WSEVLPQIDETARIIFNGKDLNVERRIAAYNPSPDLETTKPDMTLKEALKIAFGFNPN 565
DB 461 wsevlpgidettarilifngkdlnverriaavnpdpletctkpdmtlkealkiafgfnep 520
QY 566 NGNLOYQGDITREFPDPNPOQTSQNIKNQALNATNTYTVDKIKLNKKNMILIPDKRF 625
DB 521 ngnlqygdtrefdpnpoqtsqnknqlaelnactnlytvldkiklnaknmililrdrkf 580

```


PT Bacillus strain producing insecticidal protein during vegetative
 growth - used in the control of Lepidoptera and Coleoptera pests
 XX
 PS Claim 15; Page 121-124; 242pp; English.

CC Insect-specific protein VIP1a(a) (AA091239) of *Bacillus cereus* AB78
 CC shows activity against *Diabrotica* spp. pests such as the
 CC western corn rootworm. It is encoded by the VIP1a(a) gene (AA113940)
 CC isolated from a cosmid clone of AB78. VIP1a(a) can be expressed in
 CC e.g. bacterial hosts to provide biological control agents having
 CC increased activity or target range, or can be expressed in transgenic
 CC plants, esp. maize, to improve insect resistance. It is preferably
 CC expressed as a fusion protein (see also AA091245) with auxiliary protein
 CC VIP2a(a) (AA091238).

XX Sequence 884 AA:

Query Match 20.2%; Score 790.5; DB 17; Length 884;
 Best Local Similarity 26.6%; Pred. No. 9,6e-43;
 Matches 249; Conservative 136; Mismatches 295; Indels 255; Gaps 32;

```

OY 1 MKRRKVLIPLMALSTILVSS--TGNEVYQAEVK-----QENRLNSESSESSOGGLGY 51
DB 1 mkrrkvlipmlavtcltlapmflngvnavyadsktngistctknq---qkmdrkgyllgy 57
OY 52 YFSDNLFQAPMVTSTGDLSTPSELEN--IPSENYFQSAIMSGFIKVKSDDEYTPA 109
DB 58 yfsgkdf-snlcmfaptrdscllydqgtanklldkkqgygsrlwlglsksetgftfn 116
OY 110 TFSADNHVTMWDQEVINKASNSNKRRLKRGRIYQIKIYQRENPFEKGLD-----FKL 163
DB 117 lsddegalltngkllsnkgykqkvahlekylvpikleygsd--tkfnldsktlfkkl 174
OY 164 YWTDSONKKEVISDNLQPELKOKS-----NSRKRKSTSGAPVTPRDN 210
DB 175 ftdsqngqgqvgdelirpnefkkesgeflakpsklnlftqkmkreied---ctdgd 231
OY 211 GIPDSLEVEGYTVDKKRTPLSPWISNTHKGLTKRYKSPKMWSTASDPYSDFEKVTG 270
DB 232 slpdlweengylt---qnrtaevkwddsl-aaskgylkfvsnpleshtvgdpytdykaa 286
OY 271 RDKKNVSPARRHPLVAAPYIVHVDMENILSKNEDOSTNTDSEPTISKNTSTSTHRS 330
DB 287 dldlnaketlnplvaafpsvnmekvllspneils----- 323
OY 331 EVHGNAEVAANTSTRTHTSEVHGNAEVH-----AAVIDHSLSLAGEPTWAEWTMG- 380
DB 324 ---naveshststswyotegaseagiepkglisfgsvnyqhselvae---wgtstgnt 378
OY 381 --LNTADTARLANIRYVNTGTAPYIVNLPTTSLVGLKMQTLATYAKENQLSQILLAPNN 438
DB 379 sgftnaseagyllnavnyvntgcaldydvkptcsfvl-ndtlatlftaknsstalnlspe 437
OY 439 YPSKNLAIALNAODDESTPTMNYNOFLFELEKTKOLRDTDOYGINAIYNEFGNKG 498
DB 438 sypkqgqngqialtsmdfnshpflnkkqvdhllnkkpmlelngdgy---yykklkdtg 494
OY 499 RVDTSNMSEVLPQLOETPTARILIFNGKDLNLVERIRIAVNPSPDLETPTPDMTLKEALKI 558
DB 495 nlvtggenvgvlgqlakakastalivddge-rvaeekryaakdyenpedkt-psltlkdaalk 552
OY 559 AF--GFNEPENGILQYQKDIETDF--NPDQOTSNIKQLAEL-----NATNITYVLDK 609
DB 553 sypdeikelegllyllyknkpiyessvmtlydentakevtqldntgkfdvshlyav--- 609
OY 610 IRLNKKMNLILDKRHHYBRNNIIVAGADSVYKEAHREYINSTEG-----LL 657
DB 610 -kltpkmvltk-lsillydn---aesndnsigkwtlnlsvsgngkkyssnnpdanlt 664
OY 658 LN----- 659
DB 665 lntdageklnkrdyylslymkseknctgeitdgelypitctkvnnvknkdyridliah 724

```

```

OY 660 -----IDKDIRKILSGYIEIDETGL 681
DB 725 nksnpisslhktndelrlfwddistdvaaikpenldsetkqysrygikled--gi 782
OY 682 -----KEYIND-----RY-----DMLNISLRQDKTFI 705
DB 783 ldkkqglhygeflneasfnleplqnyvltkyevtysselgnpsvdtlesdklykgdtkf 842
OY 706 DPKKY--NDKLDLYISNPYKVNVAVTKENNTIN 738
DB 843 dftkyskneqglfydsqlnwdfnlnaltgdkemn 877

```

RESULT 10
 AAM19509
 ID AAM19509 standard; Protein; 884 AA.

XX AAM19509;
 XX 15-APR-1998 (first entry)

DE B. cereus VIP1a(a) protein sequence.

KW Vegetative insecticidal protein; *Bacillus cereus* strain AB78; plant;
 insect; *Sesamia nonagrioides*; maize; corn borer; toxin.

XX *Bacillus cereus*.

PN W09726339-A1.

PD 24-JUL-1997.

PF 23-DEC-1996; 96WO-EP05828.

PR 15-JAN-1996; 96GB-0000786.

PA (NOVS) NOVARTIS AG.

PI Gay PB;

DR WPI: 1997-385342/35.

DR N-PSDB: AAT733994.

PT Protecting plants against insects of the genus *Sesamia* using
 PT *Bacillus* toxic proteins - applied directly or expressed as
 PT heterologous protein by the plant, also transgenic plants expressing
 PT both Cry and VIP type toxins

PS Claim 8; Page 35-39; 168pp; English.

XX This is the amino acid sequence of the 100 kD vegetative insecticidal
 CC protein (VIP) 1a(a) from *Bacillus cereus* strain AB78. The protein can
 CC be used in a new method for protecting plants, and their progeny, against
 CC insects of the genus *Sesamia* by direct or indirect application to the
 CC plant (or seed or growing area). The protein is especially useful to
 CC protect maize plants against the Mediterranean corn borer
 CC (*S. nonagrioides*).

XX Sequence 884 AA:

Query Match 20.2%; Score 789.5; DB 18; Length 884;
 Best Local Similarity 29.4%; Pred. No. 1.1e-42;
 Matches 253; Conservative 134; Mismatches 309; Indels 165; Gaps 34;

```

OY 1 MKRRKVLIPLMALSTILVSS--TGNEVYQAEVK-----QENRLNSESSESSOGGLGY 51
DB 1 mkrrkvlipmlavtcltlapmflngvnavyadsktngistctknq---qkmdrkgyllgy 57
OY 52 YFSDNLFQAPMVTSTGDLSTPSELEN--IPSENYFQSAIMSGFIKVKSDDEYTPA 109
DB 58 yfsgkdf-snlcmfaptrdscllydqgtanklldkkqgygsrlwlglsksetgftfn 116

```

```
QY 110 TSADNHVTMWVDDQEVINKASNSKIRLEKGRLYOIKROYORENPTEKGLD-----FKL 163
DB 117 lsedegallengklisnkgkqgvnhlekylvplkleysgd--tkfndsktfkelkl 174
QY 164 YWTDSONKKEVYSSDNLQLELKOKSS-----NSRKRSTSGAPTPDRDND 210
DB 175 fklsgnqpgvgqdelrnpelfnkkesqeflakpsklnltfkmkreided---tdcdgd 231
QY 211 GIPDSLEEGYTVDKNKRFLSPWISNIHEKGLTKRKSSPEKWSASDPYSDFEYVG 270
DB 232 slpdlweengytl-----gnrlavkwddsl-askgytkfvsnpleshtvgpdydyekaar 286
QY 271 RIDKNVSPFARHPLVAAPRVHVDMENTILSKNEDOSTQNTDSETRITSKNTSRTHTS 330
DB 287 dldlsnaketfnplvaafpsvnmekvylispenils----- 323
QY 331 EVHGNAEVAHNTSTSRTHSEVHGNAEVH-----AVAIDHSLSLAGERTWAETMG-- 380
DB 324 ---nsveshstmsytnltcgasveaglgpkylsfysvnyghsetvage--wgtstgnt 378
QY 381 --LNTADTARLANIRYVNTGTAPYINVLPTTSLVLGKNOTLATIKAKENOLSOILAPNN 438
DB 379 sqfntasagylnanvynvgtgalydvkptsfvl-ndtlatitaksnstalnispge 437
QY 439 YVPSKMLAPIALMAODEFSPTTMNNOFLLEKTKQLRLDTPDQVGNATYFENGVRV 498
DB 438 sykkkqngqalatsmddfnshpiltlnkkqvdnlmnpmlletnqldg---vykikdthg 494
QY 499 RVDTSNMSEVLEPOIETTRIFENGKDLNIVERRIAAVNPSPLETETKPDMLTEKALKI 558
DB 495 nlvtgwegwvlgqikaketasllvdgge-rvaekrvaakypenedct-psltckadkl 552
QY 559 AF--GNEPFGNLOYOQCKDITFDF--NPDQOTSQNIKNOLAEL-----NATNIYVLDK 609
DB 553 sydelkelegllykxkpiyessvmltydenlaketkqjndctgkfkdvshlydv--- 609
QY 610 IKANAKNITLRDRKRFHYDNNITAVGADESVEKAHREVINSSREG-----LL 657
DB 610 -kltpkmnvtk-lsllydh---aesndnsigkwtcnlnvsgnngkkyssmnpdnl 664
QY 658 LNTD-----KDIRKILSGYVEIEDTE-----GKEVINDRYDMLN--- 693
DB 665 lntdageklknrdyalslymksekntgcetltdgelypitktkvnnkndykrldliah 724
QY 694 -----ISSLRQDGKTFIDKKRYNDKLPYLISNPNKYVNYATKENTIIINPSNGDT-S 746
DB 725 nlksnpilssih-----lktndelctlfwdldsl-tdvasikpen--ltdseikqlys 772
QY 747 TNGIK---KILIPSKGYEIG 764
DB 773 rygikledgillidckgqihyg 793

RESULT 11
AAM46712
ID AAM46712 standard; Protein; 884 AA.
XX
AC AAM46712;
XX
DT 02-JUN-1998 (first entry)
XX
DE 100 kDa VIP1A(a) toxin of Bacillus cereus.
XX
KW Vegetative insecticidal protein; VIP; expression; maize; protection;
KW plant; Ostrinia furnacalis; Asian Corn Borer; Cry toxin; VIP toxin;
KW recombinant; Bacillus thuringiensis; transgenic plant; resistance;
KW insect attack; Sesamia; maize; cereal crop.
XX
OS Bacillus cereus.
XX
PN MO9746105-A1.
XX
```

```
PD 11-DEC-1997.
XX
PF 27-MAY-1997; 97MO-EP02737.
XX
PR 06-JUN-1996; 96GB-0011777.
XX
PA (NOVS ) NOVARTIS AG.
XX
PI Hunter B, Suwantaradon K, Utdewilligen WPM;
XX
DR WPI; 1998-041787/04.
XX
DR N-PSDB; AAV16166.
XX
PT Administration of Bacillus sp. toxin protein, especially Cry or
PT vegetative insecticidal protein (VIP) protein to plants - useful for
PT protection against attack by Asian Corn Borer (Ostrinia furnacalis)
PS
PS Claim 8; Pages 43-46; 175pp; English.
CC The present sequence represents a 100 kDa vegetative insecticidal
CC protein 1A(a) (VIP1A(a)), and is derived from Bacillus cereus strain
CC AB78. The protein is used in a method for protecting plants and their
CC progeny against damage caused by Ostrinia furnacalis (Asian Corn Borer).
CC The protein is directly or indirectly applied to the plant, plant seed
CC or growing area of the plant. Cry toxins can also be used in the same
CC way, in place of VIP toxins. The Cry or VIP toxins and genes are used,
CC especially inside recombinant B. cereus or B. thuringiensis strains,
CC to produce plants protected against Asian Borer pests. Transgenic plants
CC protected against Asian Corn Borer can be used to produce seed and
CC progeny also resistant to insect attack. Plants expressing both a
CC Cry-toxin and a VIP toxin gene can also protect against Sesamia pests.
CC The method and compositions are especially used for protecting maize but
CC may also be used to protect other cereal crops against Asian Corn Borer
CC attack.
SQ
Sequence 884 AA;

Query Match 20.2%; Score 789.5; DB 19; Length 884;
Best Local Similarity 29.4%; Pred. No. 1,1e-42;
Matches 253; Conservative 134; Mismatches 309; Indels 165; Gaps 34;

QY 1 MKRRKVLIPLMALSTIIVS--TGNLEVIOAEVK-----QENRLNSESSESSQGLIGY 51
DB 1 mkmmkklasvvtctllapmflngvnavyadskntngistgkng---qkemdakqilly 57
QY 52 YFSOLFQAPNVVTSSTGDLSIPSELEN--IPSENOYFOSAIWSGFIYKKSDEYTF 109
DB 58 yfkqkdf-snlmtfaptrdstllydqqtanllldkkqgeysrlwglgsketcdffln 116
QY 110 TSADNHVTMWVDDQEVINKASNSKIRLEKGRLYOIKROYORENPTEKGLD-----FKL 163
DB 117 lsedegallengklisnkgkqgvnhlekylvplkleysgd--tkfndsktfkelkl 174
QY 164 YWTDSONKKEVYSSDNLQLELKOKSS-----NSRKRSTSGAPTPDRDND 210
DB 175 fklsgnqpgvgqdelrnpelfnkkesqeflakpsklnltfkmkreided---tdcdgd 231
QY 211 GIPDSLEEGYTVDKNKRFLSPWISNIHEKGLTKRKSSPEKWSASDPYSDFEYVG 270
DB 232 slpdlweengytl-----gnrlavkwddsl-askgytkfvsnpleshtvgpdydyekaar 286
QY 271 RIDKNVSPFARHPLVAAPRVHVDMENTILSKNEDOSTQNTDSETRITSKNTSRTHTS 330
DB 287 dldlsnaketfnplvaafpsvnmekvylispenils----- 323
QY 331 EVHGNAEVAHNTSTSRTHSEVHGNAEVH-----AVAIDHSLSLAGERTWAETMG-- 380
DB 324 ---nsveshstmsytnltcgasveaglgpkylsfysvnyghsetvage--wgtstgnt 378
QY 381 --LNTADTARLANIRYVNTGTAPYINVLPTTSLVLGKNOTLATIKAKENOLSOILAPNN 438
DB 379 sqfntasagylnanvynvgtgalydvkptsfvl-ndtlatitaksnstalnispge 437
```


AA19513
 ID AAM19513 standard; Protein; 1346 AA.
 AC AAM19513:
 XX
 XX
 DT 15-APR-1998 (first entry)
 XX
 XX B. cereus VIP1(a)/VIP2(a) fusion protein sequence.
 DE
 XX Vegetative insecticidal protein; Bacillus cereus strain AB78; plant;
 KM Insect; Sesamia nonagrioides; maize; corn borer; toxin; fusion protein.
 OS
 XX Bacillus cereus.
 XX
 PN MO9726339-A1.
 XX
 XX 24-JUL-1997.
 PD
 XX
 PF 23-DEC-1996; 96MO-EP05828.
 XX
 PR 15-JUN-1996; 96GB-0000786.
 PA
 PA (NOVS) NOVARTIS AG.
 PI
 PI Gay PB;
 DR WPI; 1997-385342/35.
 XX N-PSDB; AAT73999.
 XX
 XX
 PT Protecting plants against insects of the genus Sesamia using
 PT Bacillus toxic proteins - applied directly or expressed as
 PT heterologous protein by the plant, also transgenic plants expressing
 PT both Cry and VIP type toxins
 PS
 PS Claim 8; Page 69-75; 168pp; English.
 XX
 XX This is the amino acid sequence of a fusion protein comprising the
 CC vegetative insecticidal proteins (VIP 1A(a) and VIP2A(b) from Bacillus
 CC cereus strain AB78. The fusion protein can be used in a new method for
 CC protecting plants, and their progeny, against insects of the genus
 CC Sesamia by direct or indirect application to the plant (or seed or
 CC growing area). The protein is especially useful to protect maize plants
 CC against the Mediterranean corn borer (S. nonagrioides).
 CC
 XX
 XX Sequence 1346 AA.
 SO

Query Match 20.2%; Score 789.5; DB 18; Length 1346;
 Best Local Similarity 29.4%; Pred. No. 2,1e-42;
 Matches 253; Conservative 134; Mismatches 309; Indels 165; Gaps 34;

DB 749 dldlsnaketfnplvaafsvsvsmekvllspnenls----- 785
 QY 331 EVHGNAEVHANTSTSRHTSTSEVHGNAEVH-----AVATDHSLSLAGEETMAETMG-- 380
 DB 786 ---nsveshsstwsynlntegasaaglypkyisfgvsvnyqhseltvage--wgtsctgt 840
 QY 381 --LNTADTARLNANIRVYNTGTAPIVYVLPPTSTVLGKNOTLATIKAKENOLSOILAPNN 438
 DB 841 sqfntasaagylnanvrynnngtgalvydkpctstvl--ndiatlaltknsstalnisppe 899
 QY 439 YPFSKNLAPIALNADDFSSPTITMANYNQFLEKTRQLRLTDQVYGNATATYFENGVRV 498
 DB 900 syppkqngnlatismddfnshpflnkqvdnllnkpmmltnetqdg---vykikdth 956
 QY 499 RVDTGSWMSEVLPQIOETTARIIFNGDNLVERRIAANVPSDPLETTKPMPTKEALKI 558
 DB 957 nlvtgsewngylqikakatasilvdgge-rvaekrvaakdyenpedkt-psltlkdkalkl 1014
 QY 559 AF--GFNEPNCNLOYOGKDTIEEDF--NFDOOTSQNIKNOLAEL-----NATNITYVLDK 609
 DB 1015 sypdakeiegllyknpkplyessvmtlydentakevckqindtckfkdvshlydv--- 1071
 QY 610 IKLNAKNMLIRDRKRFHYDRNNINAVGADESYYKAEHREVINSTEG-----LL 657
 DB 1072 -klcpkmvnlk-lsilydn---aesndnsigkvtntcnivesgngkkyssmnpdanlt 1126
 QY 658 LNID-----KDRIKLISGVIETDTE-----GLKEVINDRYMLN---- 693
 DB 1127 lntdagelknrdyislymksekntqcelldagelypltkctvnnkdykrlidilah 1186
 QY 694 -----ISSLRQDQKTFIDFKKYNDKLPYISNNYKVVAVYKENTIIIPSENGDF-S 746
 DB 1187 nksnplssln-----ikndetlclfwddist--tdvasikpen--ldsetkqqls 1234
 QY 747 TNGIK---KILIFSKGGEIG 764
 DB 1235 rygikledgdlldkkggihyg 1255

RESULT 14
 AAM46723
 ID AAM46723 standard; Protein; 1346 AA.
 AC AAM46723:
 XX
 XX 02-JUN-1998 (first entry)
 DT
 XX
 XX
 DE
 XX
 XX
 KM Vegetative insecticidal protein; VIP; expression; maize; protection;
 KM plant; Ostrinia furnacalis; Asian Corn Borer; Cry toxin; VIP toxin;
 KM recombinant; Bacillus thuringiensis; transgenic plant; resistance;
 KM insect attack; Sesamia; maize; cereal crop.
 OS
 OS Synthetic.
 OS Bacillus sp.
 PN WO9746105-A1.
 XX
 PD 11-DEC-1997.
 XX
 PF 27-MAY-1997; 97WO-EP02737.
 PR
 PR 06-JUN-1996; 96GB-0011777.
 PA
 PA (NOVS) NOVARTIS AG.
 PI
 PI Hunter B, Suwantaradon K, Utdewilligen WPM;
 XX
 XX WPI; 1998-041787/04.
 DR N-PSDB; AAV16172.
 XX
 XX
 PT Administration of Bacillus sp. toxin protein, especially Cry or

PT vegetative insecticidal protein (VIP) protein to plants - useful for
 protection against attack by Asian Corn Borer (*Ostrinia furnacalis*)
 XX
 PS Claim 8; Pages 78-83; 175pp; English.
 CC The present sequence represents the fusion protein of vegetative
 CC insecticidal protein IAc (VIP1Ac) and VIP2Ac. The protein is
 CC used in a method for protecting plants and their progeny against
 CC damage caused by *Ostrinia furnacalis* (Asian Corn Borer). The
 CC protein is directly or indirectly applied to the plant, plant seed or
 CC growing area of the plant. Cry toxins can also be used in the same way,
 CC in place of VIP toxins. The Cry or VIP toxins and genes are used,
 CC especially inside recombinant *B. cereus* or *B. thuringiensis* strains, to
 CC produce plants protected against Asian Borer pests. Transgenic plants
 CC protected against Asian Corn Borer can be used to produce seed and
 CC progeny also resistant to insect attack. Plants expressing both a
 CC Cry-type and a VIP toxin gene can also protect against *Sesamia* pests.
 CC The method and compositions are especially used for protecting maize
 CC but may also be used to protect other cereal crops against Asian Corn
 CC Borer attack.
 XX
 SQ Sequence 1346 AA;
 Query Match 20.2%; Score 789.5; DB 19; Length 1346;
 Best Local Similarity 29.4%; Pred. No. 2,1e-42;
 Matches 253; Conservative 134; Mismatches 309; Indels 165; Gaps 34;
 OY 1 MKRKVLIPALMLSTIVSS--TGNLEVIQAEVK-----GENRLNSESSESSGGLG 51
 DB 463 mkmnmkklasvvtcllpmfingvnavayadsktngistctgknq--qkmdrkyllyg 519
 OY 52 YFSDLEQAPMVTSTGTDLSIPSELEN--IPSENYFQSAIWSGFTKVKSDRYTPA 109
 DB 520 yfkygxf-snlmfpdrstcllydgtlanklldkkgqygsirwglqsketgftfn 578
 OY 110 TSADNHNVTMWDQEVINKASNSNKRILEKGRLYOKIYOENPTEKELD-----FL 163
 DB 579 lsedegallnngkllsmngkqevnhlekylvpikleygsd--tkfidsktfkfekl 636
 OY 164 WYTDSONKKEVISDNQLPELKOKSS-----NSRKRKSTAGTPVPRDND 210
 DB 637 fklidngnqgqygqdelnrepefnkkesgeflakpsknlfctqkmkreided---cdtgdg 693
 OY 211 GIPDLEVEGYVVDKANKTFELSPWISNTHKKGGLTKYKSSPEKMSSTASDPYSDEKVTG 270
 DB 694 slpdlweengyftl---gnrlavkwddsl-askyylkfivsnplshstlvdydyeakaar 748
 OY 271 RIDKNVSPARHPLVAAVPIVHVDENITLSKNEOSTQNTDSEPTIKNTSTSTHTS 330
 DB 749 didlsmaketfnplyaatpssvnmekvllspnenls----- 785
 OY 331 EVHGAEVHANTSTRTSTSEVHGAEVH-----AAVAIDHSLAGERTMAETMG-- 380
 DB 786 ---nveshsstswtyltnegasveaglsfgyvsrnyghselvage--wgsstgnt 840
 OY 381 --LNTADARLANIRVYNTGTAPIYVLPPTSLVLGKNQTLATIKAKENQISQILAPNN 438
 DB 841 sqfntasagylaanvrynmvgataydvkptsfvl-ndtiatlakenslalnspge 899
 OY 439 YPSPKNIAPIALNADDPSTPTIMYNOFLELEKTKOURLTDOVYGIATINENGAV 498
 DB 900 syppkkgqngialtsmdndfshpiltlkkqvndllnmkpmmltctngdq---yykklctdng 956
 OY 499 RYDTGSNMSEVLPQIETTFARITFNGKDLNLVERRIAAVNPDPLETTTPDMLKEALKI 558
 DB 957 nlvtgngmngvltqkakatasiivddge-rvaektvaadgyenpedkt-psltlkdaiki 1014
 OY 559 AF--GFNEPNGMLYOGKIDITEFDR--NFDQOSTONIKNOQLAEL-----MATNIYTVLDK 609
 DB 1015 sydelktelegllyknpkylesswmtlydentakevtqlndtqgkfkdvshlydv--- 1071
 OY 610 IKLNKMLNLINDKRFRHIDRNNIAVGADDSVYKKAHREVINSTGB-----LL 657

DB 1072 -kltpkmnytlk-1sillydn---aesndhsigkwtctnltvsgngnkkyssnmpdanlt 1126
 OY 658 LNTD-----KDIRKITLSCYVIEEDTE-----GLKRVINDRMDLN----- 693
 DB 1127 lntdagekltnkndyylslymsekntqceitldgeilyptlctkvnvknkdykxrldilsh 1186
 OY 694 -----ISSLRDQKTFIDFKKYNDKLPYISNPYKVVVAVATKENTIIINSENGDT-S 746
 DB 1187 nksnplsslh-----ikndeditlfdwdisl-tvasklpen--lttselkqlys 1234
 OY 747 TNGIK---KILFSKKGYEIG 764
 DB 1235 rygikledgillldkkglyhyg 1255
 RESULT 15
 AAW60224
 ID AAW60224 standard; Protein; 880 AA.
 XX
 AC AAW60224;
 XX
 DT 28-SEP-1998 (first entry)
 XX
 DE Bacillus thuringiensis insecticidal toxin I77C8.
 XX
 KW Insecticide; pesticide; toxin; delta-endotoxin;
 KW biological control; lepidopteran; coleopteran.
 XX
 FT Bacillus thuringiensis strain PS177C8 (NRRL B-21867).
 FT
 FT Key Location/Qualifiers
 FH Misc-difference 253
 FT Misc-difference /note= "encoded by YTA"
 FT Misc-difference 675
 FT Misc-difference /note= "encoded by AC"
 FT Misc-difference 846
 FT /note= "encoded by RAA"
 PN WO9818932-A2.
 XX
 PD 07-MAY-1998.
 XX
 PF 30-OCT-1997; 97WO-US19804.
 XX
 PR 30-OCT-1996; 96US-0029848.
 XX
 PA (MICO) MYCOGEN CORP.
 PI Dullum CJ, Feitelson JS, Loewer D, Muller-Cohn J;
 PI Narva KE, Schmelts JL, Schnepf HE, Schwab G, Stamp L;
 PI Stockhoff BA;
 DR WPI: 1998-272226/24.
 DR N-PSDB: AAV30307.
 XX
 PT Bacillus thuringiensis isolates - used for producing pesticidal
 PT toxins and nucleotide sequences for control of lepidopterans and
 PT coleopterans
 XX
 PS Claim 5; Page 81-84; 139pp; English.
 CC This polypeptide comprises a novel soluble toxin of *Bacillus*
 CC thuringiensis (B.t.) strain PS177C8 (NRRL B-21867). The toxin
 CC belongs to a novel family of B.t. toxins that have toxicity
 CC against non-mammalian pests. Its amino acid sequence was deduced
 CC from a novel DNA fragment (see AAV30307) obtained by PCR from
 CC cellular genomic DNA of PS177C8. Disclosed and claimed are novel
 CC B.t. isolates and toxins (see AAW60218-32) that have activity against
 CC lepidopteran and/or coleopteran pests, isolated genes, probes
 CC and primers (see AAV30288-321 and AAV9734-87) useful for production
 CC of the toxins and for the identification and characterisation of
 CC these toxins, and for the identification and characterisation of

CC bacterial hosts. The invention provides 8 entirely new families of
CC toxins from B.t. isolates. The toxins have the additional ability
CC to form pores in cell membranes, and can be used to facilitate
CC entry of a second agent into a target cell.

Search completed: December 2, 2001, 13:48:31
Job time: 151 sec

XX Sequence 880 AA;

Query Match 20.1%; Score 787.5; DB 19; Length 880;

Best Local Similarity 29.5%; Pred.No.1.5e-42;

Matches 251; Conservative 199; Mismatches 311; Indels 151; Gaps 35;

```
QY 1 MKRRKVLPLMALSTIVSS--TGNLEVIQAQV-----GQNRLNESESSOGGLGY 51
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1 mkkk---lasvvtcllepmlfngvnavyadsktngisttqknq---qkemdrglly 54
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 52 YPSDLNFOAPMVVTSSTGDLSPSELEN--IPSENOYPOSAIWGFIKKSDDEYTPA 109
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 55 yfkgkdf-anlmlfaptdestllydgtankliddkqeqyslrwlglgsketgdfn 113
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 110 TSADNHVTMWVDQEVINKASNSNKRIRLEKGRLOYIKIYQRENPTFEKGD-----FKL 163
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 114 lscdeqalleingklisnkyekqkvwhlekylvpikleyqsd--tkfnidsktfkxl 171
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 164 YMTDSQNKKEVSSDNQLPELKOKSS-----NSRKRKTSAGPTVPORDND 210
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 172 fklidsqngpqvgqdelmpelnfkkesqeelaapsklnltqkmkreided---tdltdgd 228
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 211 GIPDSLEVEGYTVAVKNRKTFSLPMISNIEKKGLTKYKSSPERKWSSTADSPYSDFEKVTG 270
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 229 slpdlweengytl-----gnrlavkwddsl-askyktcfvsnplshvgtgpytdyekaar 283
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 271 RIDKNTSPEAKHPLVAPRYVHVDMEIILSKNEDOSTONTDSETRTISKNTSRTHTS 330
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 284 dldlsnaketfnpivaifsvnmekvllspneils----- 320
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 331 EVHGNAEVHANTSTSRTHTESEVGNAEVH-----AVADHSLSLAGERTWAETWG-- 380
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 321 ---nveshastnswytnntegaseagigpkyisfyvsnvghsetvage--wgtstgnt 375
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 381 --LNTADTARLNANIRVNTGTAPIYVNLPTTSLVIGKNOTLATIKAKENOLSOILAPNN 438
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 376 sgftasagylnanvynvgtgaiydvkptsfvl--nddtlatitaknsalnispge 434
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 439 YPFSKRLAPIALNADDFSTPTMTNNOPLLEKTKQLRLDQVYGNATATYFENGRLV 498
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 435 syppkkgqnglatismddfnspitlnkkqvdnlmnpmmletnqtdg---vykkikdthg 491
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 499 RVDTGSMNSEVFLPQIOETTFARIIFNGKDLNIVERRIAANVPSDPLETTKPPDMLKEALKI 558
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 492 nlvtgqewngviqklakaktasiliyddge--rvaekryaakdyenpedkt--psltlkdaiki 549
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 559 AF--GENEPNGNLQYOGKDIPEDF--NPDQOTSNIKOLAEL----NATNIYTVLDR 609
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 550 syppdelkelegllyknkpiyessvmtlyidentakavtkglndtltgkfkdvshlydy--- 606
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 610 IKUNAKNLIIRKREHYDRNNIYAGADESVYEAHREVINSSTEG-----LL 657
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 607 -kltpkmnytlk--lsillydn--aesndnsigkwtlnlsvsgnngkkyssnmpdanlt 661
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 658 LNIID-----KDIRKILSGYIVEIEDTE-----GLKEVINDRYDMLNI--SS 696
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 662 lntdeqeklnkryislymksekntgcetltdgelypitlctvnnkdnkykldilahn 721
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 697 LRODGTFTIDFKRYNDKLPLYISNPNYKVVAVTKENTTIINPSENGDT--STNGIK--K 752
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 722 lksnpisslnktc-ndetltfwddisi--tdvasikpen--ltdselkqiystrygikledg 777
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 753 ILFSSKKGYEIG 764
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 778 lildkkqglnhyg 789
  |||: : : | : | : | : | : | : | : | : | : | : | : | : | : |
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:51:36 ; Search time 72.22 Seconds
(without alignments)
588.554 Million cell updates/sec

Title: US-09-747-521-4_COPY_178_735

Perfect score: 2871

Sequence: 1 DNQLPELKKKSSNRKRKRS.....LYISNPKNKVAVATKENT 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2758	96.1	764	2	I39934 protective antigen
2	715.5	24.9	875	2	I40862 Iota toxin compone
3	211.5	7.4	192	2	I39933 cryptic protein -
4	211.5	7.4	204	2	G59104 hypothetical prote
5	181.5	6.3	2178	2	S55805 alpha-toxin - Clos
6	164.5	5.7	2529	2	B64635 toxin-like outer m
7	158	5.5	4688	2	F82885 hypothetical prote
8	157	5.5	1308	2	E71622 probable membrane
9	150.5	5.2	2269	2	T28677 thobtry protein -
10	150.5	5.2	2399	2	H71879 toxin-like outer m
11	150	5.2	1127	2	T28317 ORF MSV156 hypothe
12	150	5.2	1230	2	S56850 SMCI protein homol
13	150	5.2	4919	2	T31105 hypothetical prote
14	149	5.2	1193	2	S68218 botulinum neurotox
15	148	5.2	810	2	D86818 hypothetical prote
16	147	5.1	1169	2	T18423 hypothetical prote
17	147	5.1	1939	2	T18372 repeat organelle
18	145	5.1	2401	2	T28676 thobtry protein -
19	144	5.0	2523	2	T18704 hypothetical prote
20	143	5.0	2340	2	B71704 cell surface anti
21	142.5	5.0	1072	2	A86827 hypothetical prote
22	142.5	5.0	2526	2	T20531 hypothetical prote
23	142.5	5.0	6658	2	T13931 projectin - fruit
24	142	4.9	1595	2	T31082 endo-1,4-beta-xyla
25	141.5	4.9	4152	2	T31102 filamentous hemag
26	141	4.9	1658	2	S55101 hypothetical prote
27	140.5	4.9	2539	2	B71619 hypothetical prote
28	139	4.8	1358	2	A29360 SIR4 protein - yea
29	138.5	4.8	1252	2	B42771 reticulocyte-bind

30	138.5	4.8	1928	2	S46773 myosin heavy chain
31	138	4.8	1033	2	T37715 actin-interacting
32	137	4.8	1946	2	JC6032 lactocep (EC 3.4
33	136.5	4.8	805	2	G82884 hypothetical prote
34	136.5	4.8	821	2	S67087 hypothetical prote
35	136	4.7	1465	2	S31262 Tyb protein - yeas
36	136	4.7	1790	2	S67593 transport protein
37	136	4.7	1803	2	S56894 Tyb protein - yeas
38	135.5	4.7	3724	2	T18427 hypothetical prote
39	135	4.7	1802	2	S52611 Tyb protein - yeas
40	134.5	4.7	624	2	PC6003 surface membrane p
41	134.5	4.7	1558	2	B71603 RESA-H3 antigen p
42	134	4.7	753	2	S48267 probable membrane
43	134	4.7	1570	2	T18272 1-phosphatidylinos
44	134	4.7	1711	2	T18429 hypothetical prote
45	134	4.7	2722	2	T20532 hypothetical prote

ALIGNMENTS

RESULT 1
I39934
Protective antigen precursor - Bacillus anthracis plasmid
C:Species: Bacillus anthracis
C:Date: 19-Jul-1996 #sequence, revision 19-Jul-1996 #text_change 01-Dec-2000
C:Accession: I39934; S69160; F59104
R:Welkos, S.L.; Lowe, J.R.; Eden-McCutchan, F.; Vodkin, M.; Leppla, S.H.; Schmidt, J.
Gene 69, 287-300, 1988
A:Title: Sequence and analysis of the DNA encoding protective antigen of Bacillus ant
A:Reference number: I39933; MWID:8912073
A:Accession: I39934
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Residues: 1-764 <RES>
A:Molecule type: DNA
A:Residues: 1-313, 'O', 315-764 <OKI>
R:Fieldman, T.C.; Gordon, V.M.; Leppla, S.H.; Klimpel, K.R.; Birch, N.P.; Lob, Y.P.
Arch. Biochem. Biophys. 316, 5-13, 1995
A:Title: In vitro processing of anthrax toxin protective antigen by recombinant PC1 (
A:Reference number: S69160; MWID:95142670
A:Accession: S69160
A:Molecule type: protein
A:Residues: 197-202 <FRT>
R:Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koeh
J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of pXOI, the large Bacillus anthracis plasmid harb
A:Reference number: A59091; MWID:99445483
A:Accession: F59104
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-313, 'O', 315-764 <OKI>
A:Cross-references: GB:AF065404; NID:94894216; PIDN:AAD32414.1; PID:94894326
A:Experimental source: strain Sterne
A:Note: similar to anthrax toxin moiety, protective antigen, pagA formerly pag, plasm
C:Genetics:
A:Gene: pXOI-110
A:Genome: plasmid
C:Function:
A:Description: three component exotoxin; protective antigen binds to receptors on the
Y active components edema factor or lethal factor; the complex is internalized by rec
C:Keywords: exotoxin
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-196/Domain: propeptide #status predicted <PRO>
F:197-202/Product: protective antigen #status experimental <MAT>

Query Match 96.1%; Score 2758; DB 2; Length 764;
Best Local Similarity 96.8%; Pred. No. 4.7e-144;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;
Oy 1 DNQLPELKKKSSNRKRKRSAGPTVPDRNDGIPSLVEGYTVVKKRFFLSWIS 60
|||||
Db 178 DNQLPELKKKSSNRKRKRSAGPTVPDRNDGIPSLVEGYTVVKKRFFLSWIS 237

Query Match	6.38;	Score 181.5;	DB 2;	Length 2178;
-------------	-------	--------------	-------	--------------

```
0Y      120 NIIISK--EODSTONTSEFTIKSNSTSTTHSEVHNAEVAHNSTSRTHSEVHG 177
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DbB     825 NIYLINNETGSGVNSDGGCANITFKASDNITMDGLNDAE-----IYTKMIQT 875
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y      178 MAEVHAVIDHSL-----SLAGERTW-----AETMGLTAD-----TAR 211
```

Db 876 GASQSHYAFDALNINISYNSFSDDMTWGFSPSAKNISFNSRGCTNPGSSVISAN 935
Oy 212 LMANIRYVNT--GTAPIYVLPPTSLVLGKNQTLATI-----KAKENQSLIAPANNYP 264
Db 936 ATNLSIFINSRLNGCAVYN--LQANSLIFNNQAVFVVLVSRGNSNNATTOGLGNFTL 994
Oy 265 S-----KLAIAL--NAQDDSSPTITMANTNOELEKTKQRLDLDQVYG 309
Db 995 SSQSLNFGNDTTLQNNANITLGNKSQAAPKNS--LTLDDNSNLSLDNOSVLNANNNTSAFN 1053
Oy 310 NIATYFENGGRVVDGSMSEVLPIQIETTARIENGKDLN--VERRIAAYVPSDPLET 368
Db 1054 NQASLNTIYGS-----QATENSLEFNGGTLSLMSSKKNASNASFSNNT 1097
Oy 369 TRPDMTLKALIAEGFNPNGLQYQKDIIEF-----DENFDQOTSQNIKNOLA 419
Db 1098 T---INLDDSVLASNTSSLNANINFGASQADFGNTIIDIASFPDSSASLNFNNLLA 1154
Oy 420 -----ELNATITVYLDKIKLNAKMNILLRDKRF- 448
Db 1155 NCALNFGYTPSLTYALMSVSGQFVLGNNGDINLSDI--NIFDNITKSVTYNILLNAQKGIT 1213
Oy 449 -----HYDRNNTAVGA-----DESVYKFAHR-----EVIN- 473
Db 1214 GISGANGYEKILFYGMKIQMNTATYSDNNNTQTSFNPPLMSQIIOISIKNGDITTEVLNN 1273
Oy 474 -STEGGLINIDKDI-----RKILSGYIYEIEDTEGLEKVINDRYDMLNSSL----- 520
Db 1274 PMSASNTITINIAPELYNAYQASKNPFGYSYSDNQ-----GYTYLTSNIGLFPKGS 1328
Oy 521 --RODGKTFIDKKYNDKPLX---ISNPNYVNVYAVTKENT 558
Db 1329 QTPQAPGYTSPNPQLSSLNLYNKGFSSENLK--TLGLGISONS 1370

RESULT 7
P82885
hypothetical protein U0482 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: F82885
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Casseil, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
A:Reference number: A82870
A:Accession: F82885
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4688 <GLA>
A:Cross-references: GB:AE002145; GB:AF222894; NID:96899476; PIDN:AAF30894.1; GSPDB:GNO01
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: U0482
A:Genetic code: SGC3

Query Match 5.5%; Score 158; DB 2; Length 4688;
Best Local Similarity 21.2%; Pred. No. 4.1;
Matches 138; Conservative 107; Mismatches 270; Indels 136; Gaps 32;

Oy 1 DNLOQLPELKOKSSN---SRKRSTSAGTPVDRD-----NDGIFD 37
Db 3803 DNLN--PETRYKLENIELSKPLKTHNLVSINDKENISLITETGNPVLKIOTONDITND 3861
Oy 38 SLEVEGYVDVYENKRRFLSPWISNIHEKGLTKYKSSPEKSTASDPYDFEKVYGRGD 97
Db 3862 TQQTINVLISGVNSK--YNGRQIKVYKKNNNVITYESS---LITLQKGKNDYOLLNLNS 3917
Oy 98 NVSPARHPPLVAAYPIVHV-----DME-----NIIISKNEOSTONTDSE-----TR 139
Db 3918 N-----REVREFEIKIINHISNTNPFDELEKLGVSNTFTTQTKNTTYVQNDSSATIVGR 3972

Oy 140 TISKNTSTSRHTSEVHGAEV---HANTSTSRHTSEVHGAEVAAVADHSLSLAGER 196
Db 3973 GVNFNFKI--KSEDKLIENNQQVAVAPAKKETIHDITWLTQYRPLKDVSDFK-----EG 4026
Oy 197 TMAETGLMT---ADTARLANIRYVNTGTAPIYVLPPTSLVLGKNQTLA-----TIK 247
Db 4027 TMAHDLNSVNFKEETTYLVKIQFVNKPKKANNINSENNVILDNITSINSNEFTTK 4086
Oy 248 AKENQSLIAPANNYPKSNLAPIALN--AODDESSPTITMANTNOELEKTKQRLD 304
Db 4087 VGDHKLINTSSNNVNTFNOTINFTLSGVKKSVMGKKIKLSTKSDNTSSIHTEVLISS 4146
Oy 305 DOVYGNIAFYENFNGR-----VRDGSNMSEVLPIQIETTARIENGKDLNVERRIA 358
Db 4147 NKTQYNILLNNLKRNTTYTLIDVKLIDNNNVSDPEKRGMLITNSFITRPSAINVLEEI 4206
Oy 359 AVNPSDPLET-----KPDMTLKEA-----LKIAPFNENGSLQYQKDIIEF 402
Db 4207 SNRASNLSKSTIIKINLNDPVDVLRDQDATTYYGNNKQAMGFTYVSGNIKYLTLATVDL 4266
Oy 403 DENFDQOTSQNIK--NOLAEINATNTYTVLDKIKLNAKMNILLRDKRFHYD-----RNN 454
Db 4267 NFN--DKVNIIVNISPNKKPSIAEN-----IGDKSNII-----YNNDSIPKLEINND 4313
Oy 455 IAVGA---DESVYKFAHREYVINSSTEGLLNIDKDIRKILSGYIYEIEDTEGLEKVIN- 510
Db 4314 IIVNGPINKIEIYKKNANOK--NNIDVDLQIINPKIAHMLR--FIKFKSTN-----NDI 4364
Oy 511 -RYDMLNSSL--RODGKTFIDKKYNDKPLXISNPNYK--VNVYAVTKENT 558
Db 4365 IETNVINGSSLVANNQDKTISRFTLNNLK-----ANKLYSLVDVYLVVNNNS 4410

RESULT 8
E71622
probable membrane associated protein PFB0125c - malaria parasite (Plasmodium falcipar
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: E71622
R:Gardner, M.J.; Telieps, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743
A:Accession: E71622
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1308 <GAR>
A:Cross-references: GB:AE001374; GB:AE001362; NID:93845100; PIDN:AACT1815.1; PID:9384
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0125c

Query Match 5.5%; Score 157; DB 2; Length 1308;
Best Local Similarity 19.9%; Pred. No. 0.72;
Matches 130; Conservative 95; Mismatches 219; Indels 210; Gaps 31;

Oy 47 DYKNNRTFLSPWISNIHEKGLTKYKSSPEKSTASDPY----- 85
Db 42 DYKEDVTL-----NIDKKKNVEEYKKNKDYKNEDEFFVFDKEIKLKKIKEBOCN 96
Oy 86 ---SDFEKYTGRI--DKNVSPARHPPLVAAYPIVHVDMENIILSKNEOSTONTDSETRT 140
Db 97 MKNEFINEKGYIILNDVENS--TINNITSLNDIILHSSDNCVCTSYNIYPSNNNNNN-- 153
Oy 141 ISKNTSTSTRTHTSEVHGAEVANSTSTRTHTSEVHGAEVAAVADHSLSLAGE----- 195
Db 154 -----NNNVNHSNNS-----NIFVDSHMEHDDITDEFKID 187
Oy 196 RTMAETMGLTADTARLANIRYVNTGTAPIYVLPPTSLVLGKNQTLATIAKKNQISO 255
Db 188 QTNSEFPQNFISFENKKNVNEBELMKKHTDNI--NI---CDKIIDK-----KKNCNTLSD 236

QY 206 TADT-----ALRLANTRY-----VNTGAPLYNYL-----PTSLVLGKNOTLA 244
 : : : : : | | | | | : : : : : | | : : : : :
 Db 807 AASNSLSTFSSRLNGALITNLQANSILFINNTQAVFNWLYLSGTSFNATITQULGTFSTLS 866
 QY 245 TIKAKENQLSLIAPNYYPSKNIAPIAL--NAODFFSPTITMNYNOFLELEKTKQURL 302
 : : : : : | | | | | : : : : : | | : : : : :
 Db 867 S-----QSLINFGDPFTLLQNNANITIGKNSQAFAFKNS-LTLDNNSNLSDMQSYLNA 917
 QY 303 DDDYYGNATATYFENGRRRVDTGSMWSEVLPIQETATIIINGKDLNVERIAAVNP 362
 : : : : : | | | | | : : : : : | | : : : : :
 Db 918 NQTSAPANNQASINLYNGS-----QAASFSLFENGTLSTL--NANSKRLNA 959
 QY 363 SDPLETKRDMTLKEALKIAFGFENEENGLYOOGKIDTER-----DENFDQOQSON 413
 : : : : : | | | | | : : : : : | | : : : : :
 Db 960 SSASFSNNNTINLDSDVLANANTSSILANINIFQASQADPDGGMTTIDTASFNFDSASSLN 1019
 QY 414 IKNOLA-----ELNATNIYTVLDKIKLAKNNILI 443
 : : : : : | | | | | : : : : : | | : : : : :
 Db 1020 FNNLTANGALNFGYAPSLTKALMNSGQFVLGNNDINLSDI-NIPDNITKSVTYNINL 1078
 QY 444 RDKRF-----HYDRNNIIVGA-----DESVYKEAHR----- 469
 Db 1079 AOKGTTIGSANGYEKILFYGMKIQNATYSDDNNITQWSTINPLNSQIIQESIKNGDLT 1138
 QY 470 -EVIN--SSTEGILLNID-----KDIRKILSGIYEIEPTGELKEVINDRYDMLNISTSL 520
 : : : : : | | | | | : : : : : | | : : : : :
 Db 1139 IEVLNPNNSASNTIFIAPELWYQSKQMPGTGYSDYSDNQ-----GRYILTSNITKGL 1193
 QY 521 -----RQDKGTFTDFKKYNDKLLY---ISNPYKVVAVYAVTKENT 558
 : : : : : | | | | | : : : : : | | : : : : :
 Db 1194 FTRPKSGQTPQTPGTYSFNPQPLNSLIYNGKFSSEMLK-TLLGILSLONS 1241
 : : : : : | | | | | : : : : : | | : : : : :
 RESULT 11
 T28317
 ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
 C:Species: Melanoplus sanguinipes entomopoxvirus
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
 C:Accession: T28317
 R:Alfonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
 J. Virol. 73, 533-552, 1999
 A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
 A:Reference number: Z20484; MUID:99102612
 A:Accession: T28317
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1127 <AFO>
 C:Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97677.1; PID:g4049717
 C:Genetics:
 I:Note: MSV156

```

Query Match          5.2% Score 150; DB 2; Length 1127;
Best Local Similarity 20.8%; Pred. No.1.4;
Matches 116; Conservative 103; Mismatches 226; Indels 114; Gaps 28;

QY      44 YTVVAKNRKRTFLSPISNINHERKGTLTKYSSDEKSTADSPYSDEEK-VTGR---ID---96
       | : : : : : | : : : : : | : : : : : |
Db      117 YDLDEEKKD--KELVINECKNAVDKIDIK---NNVNHNHSDNTITIGGETLIDIIN 171

QY      97 ---KVVSPEARHPPLVAAPYIVH---VDMENI-----ILSKNEODSTONTSETPTISKMT 145
       | : : : : : | : : : : : | : : : : : |
Db      172 KLLKLVSSDEKOLIQIYYKNINKKEIEFENNINOVEIKRKODEKLKLDESKEKEPIFK- 230

QY      146 STSRHTTEGVHGNAAEVHANITSRHTTSYVHGNAAEVHAAYADHSLSLAGERTWAEEMGIN 205
       | : : : : : | : : : : : | : : : : : |
Db      231 -----DELINTDKOEELIKLKNDXEIENPNIIDEOKKLLDOINS-----KIN 273

QY      206 TADTARLANAIRYV-NTGTAPIYVLPFTSYLGKNQOTATIGAKENOLSOLAPRNYPYR 264
       | : : : : : | : : : : : | : : : : : |
Db      274 T-----LNENIKGVANLYTETETKKNKSIGNQLNELNKDDSTIKSLDEKOKLLDELDKINNIT 328

```

```

QY 265 S-----KNLAPILANQDFFSHPIMYANQGLEKTYKQRLDLD--OVYCNATAY 314
      | : : : : : | : : : : : | : : : : : | : : : : : |
Db 329 SLVKSNTKRTITNOQLLESSTLDFNNANITIN-----EL-KSKIKLFDNDIKLANNIDTE 382
      | : : : : : | : : : : : | : : : : : | : : : : : |
QY 315 NEFNGRVAVDTGSNMSEVLPOLQIETTARIIFNGKDLNL--VERRIAAVNPSPDLETTYKPD 372
      | : : : : : | : : : : : | : : : : : | : : : : : |
Db 383 --QNNKI-LTDFPNNSSTRIFEXKRLDTEYKRIIDIKNNNLQKLESBYKKIDEGEYKKNIN 439
      | : : : : : | : : : : : | : : : : : | : : : : : |
QY 373 MTLKEALKIARFENFNGNLQYQKDIETFEFENEDQOTS-----ONIKNQLAELNATNI 426
      | : : : : : | : : : : : | : : : : : | : : : : : |
Db 440 KEYNDITEL-----KNNNLQ-----KLEENKKRIDQTEYKRIKINYEYNDITELKNNNL 489
      | : : : : : | : : : : : | : : : : : | : : : : : |
QY 427 YTVLADTK-LMAKKNILIRKRFH---YDRNNIANGADESVYKEAHREYINSTE----- 477
      | : : : : : | : : : : : | : : : : : | : : : : : |
Db 490 QKLEENKKNINDKLTIKLNDIESNTEFLKNTL-----SDFKDKSREIAKLNTEYOLR 543
      | : : : : : | : : : : : | : : : : : | : : : : : |
QY 478 -GILLNIDK--DIRKILSGYIVIEDTEGELAKYEYNDRYMLNINSLSROGKFFIDPKRYN 534
      | : : : : : | : : : : : | : : : : : | : : : : : |
Db 544 KDLLENINKKNEIMLKISDNKLSSLEOLYDSKKNILDID-----KIYNSLKEKN 592
      | : : : : : | : : : : : | : : : : : | : : : : : |
QY 535 DKLEPLYISNPYKVNAYV 553
      | : : : : : | : : : : : | : : : : : | : : : : : |
Db 593 DKIDEIFSNIE-KFDIYNV 610
      | : : : : : | : : : : : | : : : : : | : : : : : |

```

RESULT	12
S56850	

SMC1 protein homolog YJL074c - yeast (*Saccharomyces cerevisiae*)
N: Alternate names: probable membrane protein YJL074c; protein J1049

Cispecies: saccharomyces cerevisiae
C:Date: 05-May-1995 #sequence_revision 08-Sep-1995
C:Date: 05-May-1995 #text_change 29-Sep-1999

C/Accession: 550650; 55113/
R/Rose, M.; Koetter, P.; Entian, K.D.
Submitted to the Protein Sequence Database
September 1995

Submitted to the Protein Sequence Database, September 1993
A/Reference number: S56848
A/Accession: C56850

A;Accession: 536630
A;Molecule type: DNA
A;Date: 1-1990

A;Residues: 1-1230 <ROS>
A;Cross-references: EMBL

submitted to the EMBL Data Library, June 1995

A;Accession: S57737

A;Molecule type: DNA
A;Residues: 1-1230 <

Cross-references: EMBL:AB8851; NID:g895892; PID:g895899
Genetics: [Genetics](#)

A:Gene: SGD:SMC3
A:Cross-references: SGD:S0003610; MIPS:YJL074C

C; Superfamily: hypothetical protein YJL074c

C;Keywords: transmembrane protein

Query Match	5.2%	Score 150;	DB 2;	Length 1230;
Best Local Similarity	20.4%	Pred. No. 1.6;		
Matches 138;	Conservative 102;	Mismatches 255;	Indels 182;	Gaps 28;

[illegible]

Db 620 KKENISMPRIIESYEIPNDMLGLPLNDLNEKLFNITSKNTATPKKIYYNFDOMWTQYYSQ 679
QY 233 --TSLVLKNOCTLA--TIKAKENQLSOILAPNNYPSKNIAPALNA----QDDFSSTP 283
Db 680 YEDLICMAKRSVLAQETLIKRIIOKKLSYLGINSNISSDNLALMLTTNTLRDISNSQ 739
QY 284 ITMN-----YNOFLEKTKOLRLTDQYVG--NATYFENGRRV 323
Db 740 IAMNVDSFLNNAALCVESNITPKFISF-----MEOCINNINIKTKEF---IOK 786
QY 324 DTGSNMSEVLPOIOETTARIIFNGKDLNVERRIAANVSPDLFTTKPDMTKEAL---- 379
Db 787 CTNINEDERKOLINON-----VENSIDFEELN-----IQMKSLSFSEALILIKETWYE 837
QY 380 KIAFGNEPNNGN-----LOYGKDI--TEPDFNFD-----QOTSONIKQLAEL 421
Db 838 LVLYAFKPEGNVIGDASGKNTSIEY--SKDIGLYGINSDALYNGSISFSNDFFEN 896
QY 422 NATNIYTV-----LDKIKLAKM-----NIIIR 444
Db 897 GLTNSFSIYFWLRNLGKDTIKSLKLGSKEDNCWEIYFQDTGLVENMTDSNGENKIYLS 956
QY 445 D---KRFHY-----DR--NNIAGADESVY--KEAHREYINSSTEGLLNIDKIKIISG 493
Db 957 DVSNSMSWHYITISVDRLKEQLIFIDNULVANESIKETIINITYSSNISLSENNPSYIEG 1016
QY 494 YIVEIEDTEGLKEVINDRYDMLNISSLRQDKTFIDFK-----YNDKLPYISNPN 545
Db 1017 LTIILKPTTS--QEVLSNFEVLYNSYIRDSNERLEYNNTQYLYNVVSDPKICEVKQNN 1075
QY 546 YKVNYYAATKENT 558
Db 1076 --NIY-LTINNT 1084

RESULT 15
D86818
hypothetical protein ypl1 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
C:Accession: D86818
R:Boletín, A.; Wincker, P.; Manger, S.; Jallion, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. In press, 2001
A:Title: The complete genome sequence of the lactic acid bacterium.
A:Reference number: A86625
A:Accession: D86818
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-810 <STO>
A:Cross-references: GB:AE005176; MTD:g12724549; PIDN:AAK05646.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: ypl1

Query Match 5.2%; Score 148; DB 2; Length 810;
Best Local Similarity 21.1%; Pred. No. 1.1;
Matches 115; Conservative 80; Mismatches 203; Indels 146; Gaps 25;

Db 200 LDGINITTKANNLYPNNFLTPDGLHMSDNTQYIVY-----KGAATG 240
QY 310 NIATYNE--ENGRRVVDY-----GSNMSEVLPOIOETTARIIFNGKDLNVERRIAANVPS 363
Db 241 QISSENFGLTGPTTLATSYTMNMGDSGRITNVGRTLA-----GTNLDLIYKIVISTDESS 295
QY 364 --DPLETTKPKMTLKEALK-----IAFGENEPNGLOYOGKDIITEPDFNF-- 406
Db 296 WQAPKESTS--DCPIGLAFTGEONIANSDGNSIVALYGANNVNLYQ---IVVHNTNFOI 351
QY 407 -----DOOTSONIKNOLAEIATNIYTVLDKIKLAKMNLIRDKRFRHORN--NI 455
Db 352 PYLASFTITDIDMAGVNTLA-----NLTVIRK-----TTNLATDSSIIYDTTIPDT 401
QY 456 AVGADESYYKEAHREYINSSTEGLLNIDKD-----IRKILSGYIE 497
Db 402 DLNGASLPYGGYLGV-----GFLSNFDYDFYSPAPARSGDSYQSGGVHVDLFGSALQ 455
QY 498 IEDTEGLKEVIN--DRYDMLNISSLRQDKTFIDFKKYDKLPYISN--PNYKVN--VYAVT 554
Db 456 AHLTQVVRDIYLYNLYDADENNQLIVPOHOETWFPNVSMMNP--VSNPHYQYGMQNOTT 513
QY 555 KENT 558
Db 514 KQNT 517

Search completed: December 2, 2001, 13:51:43
Job time: 338 sec

QY 107 LVAAAPPI-VHVMENIILSKNEDOSTON-----TDSEFTISKMT 145
Db 28 LVTSIPLAVKADENSAVTPPNQTEAVTNPVNSSLAASDSTEGNTGSSQAEYRTKEANS 87
QY 146 STSRHTSEVHGNAEVHANTSTSRHTSEVHGNAEVHAVAIDHSLSLAGEPTAETMGLN 205
Db 88 LTSQASTLL---SASSVSSSTSSQOTSSV--EASSNSSEVSASLSKASTKAPSVLP- 141
QY 206 TADTARLANIRYVNTGAPYVNLPTTSLVL---GKNQTLATIKAKENQLSOILAP-- 259
Db 142 --DSSKANVTYIASSLSAGTIIIPALTNQSIQANGODPVISETATGVSANDITGPAT 199
QY 260 -----NNYPSKNIAPALNAODFSSTPTTMNTNOFLELEKTKOLRLDQYVG 309

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:48:42 : Search time 112.1 Seconds
(without alignments)
368.715 Million cell updates/sec

Title: US-09-747-521-4_COPY_178_735

Perfect score: 2871
Sequence: 1 DNLQPELKOKSSNSRRKKRS.....LYISNPKNVAVYATKENT 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_1101: *
1: /SID88/gcgcdata/geneseq/geneseq/AA1980.DAT: *
2: /SID88/gcgcdata/geneseq/geneseq/AA1981.DAT: *
3: /SID88/gcgcdata/geneseq/geneseq/AA1982.DAT: *
4: /SID88/gcgcdata/geneseq/geneseq/AA1983.DAT: *
5: /SID88/gcgcdata/geneseq/geneseq/AA1984.DAT: *
6: /SID88/gcgcdata/geneseq/geneseq/AA1985.DAT: *
7: /SID88/gcgcdata/geneseq/geneseq/AA1986.DAT: *
8: /SID88/gcgcdata/geneseq/geneseq/AA1987.DAT: *
9: /SID88/gcgcdata/geneseq/geneseq/AA1988.DAT: *
10: /SID88/gcgcdata/geneseq/geneseq/AA1989.DAT: *
11: /SID88/gcgcdata/geneseq/geneseq/AA1990.DAT: *
12: /SID88/gcgcdata/geneseq/geneseq/AA1991.DAT: *
13: /SID88/gcgcdata/geneseq/geneseq/AA1992.DAT: *
14: /SID88/gcgcdata/geneseq/geneseq/AA1993.DAT: *
15: /SID88/gcgcdata/geneseq/geneseq/AA1994.DAT: *
16: /SID88/gcgcdata/geneseq/geneseq/AA1995.DAT: *
17: /SID88/gcgcdata/geneseq/geneseq/AA1996.DAT: *
18: /SID88/gcgcdata/geneseq/geneseq/AA1997.DAT: *
19: /SID88/gcgcdata/geneseq/geneseq/AA1998.DAT: *
20: /SID88/gcgcdata/geneseq/geneseq/AA1999.DAT: *
21: /SID88/gcgcdata/geneseq/geneseq/AA2000.DAT: *
22: /SID88/gcgcdata/geneseq/geneseq/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2758	96.1	735	15	AA60179
2	2758	96.1	736	21	AA156959
3	2758	96.1	763	21	AA156960
4	2758	96.1	764	22	AA156958
5	2758	96.1	764	22	AA156958
6	2758	96.1	764	22	AA156958
7	2664	92.8	569	21	AA156961
8	2581	89.9	719	15	AA60193
9	593	20.7	884	17	AA601239
10	591.5	20.6	880	19	AA60224
11	591	20.6	881	21	AA159277

12	589	20.5	884	18	AA19509	B. cereus VIP1(a)
13	589	20.5	884	19	AA19512	100 kDa VIP1(a) t
14	589	20.5	1346	17	AA191245	VIP2(a) and VIP1A
15	589	20.5	1346	18	AA191245	B. cereus VIP1(a)
16	589	20.5	1346	19	AA191245	VIP1(a)/VIP2(a)
17	588	20.5	852	17	AA191246	VIP1(a) protein w
18	588	20.5	852	18	AA191246	Maize optimised-B.
19	588	20.5	852	19	AA191246	Maize optimised-VI
20	588	20.5	1338	17	AA191247	VIP2(a)-VIP1(a)
21	588	20.5	1338	18	AA191247	Maize optimised-B.
22	588	20.5	1338	19	AA191247	VIP2(a)/VIP1(a)
23	584.5	20.4	884	15	AA19520	Bacillus cereus 10
24	584.5	20.3	667	17	AA191240	B. cereus VIP1 pro
25	583	20.3	860	21	AA195282	MIS-8 toxin from B
26	582.5	20.3	784	21	AA195282	MIS toxin of Bacil
27	580	20.2	667	18	AA19510	B. cereus 80 kD VI
28	580	20.2	667	19	AA19510	80 kDa VIP1(a) to
29	578	20.1	667	15	AA19513	Bacillus cereus 80
30	572.5	19.9	834	17	AA191242	B. thuringiensis V
31	572.5	19.9	834	18	AA191242	B. thuringiensis V
32	572.5	19.9	834	19	AA191242	Vegetative insect
33	411	14.3	425	19	AA19522	Bacillus thuringie
34	318	11.1	357	19	AA19523	Bacillus thuringie
35	315	11.0	357	19	AA19523	Bacillus thuringie
36	307	10.7	357	19	AA19523	Bacillus thuringie
37	303	10.6	357	19	AA19523	Bacillus thuringie
38	291	10.1	357	19	AA19523	Bacillus thuringie
39	288	10.0	327	19	AA19520	Bacillus thuringie
40	278	9.7	327	19	AA19520	Bacillus thuringie
41	256.5	8.9	333	19	AA19523	Bacillus thuringie
42	252	8.8	347	21	AA195276	Toxin from B. thur
43	252	8.8	348	19	AA195229	Bacillus thuringie
44	252	8.8	348	19	AA195229	Bacillus thuringie
45	252	8.8	348	19	AA195229	Bacillus thuringie

ALIGNMENTS

RESULT 1	
AA60179	standard; Protein: 735 AA.
XX	
AC	AA60179;
XX	
DT	03-APR-1995 (first entry)
XX	
DE	Protective antigen of Bacillus anthracis.
XX	
KW	Anthrax; Bacillus anthracis; fusion protein; protective antigen;
KW	protective antigen; cell killing; targeting; pathogen;
KW	intracellular; HIV; human immunodeficiency virus; toxin.
XX	
OS	Bacillus anthracis.
XX	
PN	WO9418332-A.
XX	
PD	18-AUG-1994.
XX	
PF	14-FEB-1994; 94WO-US01624.
XX	
PR	12-FEB-1993; 93US-0021601.
XX	
PR	25-JUN-1993; 93US-0082849.
XX	
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.
XX	
PI	Arora N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;
XX	
DR	WPI: 1994-279753/34.
XX	
DR	N-PSDB: AAQ70180.
XX	
PT	Nucleic acid encoding anthrax toxin fusion protein - useful for
	targeting toxin to specific cells, eg for killing tumour cells

or HIV-infected cells

Disclosure: Page 81-83; 124pp; English.

The sequence encoding the protective antigen of *Bacillus anthracis* may be used in the construction of a nucleic acid which encodes a fusion protein comprising the anthrax protective antigen binding domain of the native anthrax lethal factor and a sequence encoding an activity inducing domain of a second protein. The fusion proteins are useful for the specific killing of tumour cells or the killing of cells infected with intracellular pathogens, especially HIV.

Sequence 735 AA:

Query Match 96.1%; Score 2758; DB 15; Length 735;
Best Local Similarity 96.8%; Pred. No. 7,4e-188;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

```
OY 1 DNLQLPELKQKSSNRKRSTAGPTVPDRNDGIPDSLEVEGYVVDVKNKRTFLSPWIS 60
    |||||||
DB 149 dnlqlpelkqkssnrkrstagsptvpdrndgipdslevegylvdvknkrtflspwis 208
OY 61 NHEKKGLTKYKSSPEKKSTASDPYSDEPKYGRIDKNVSPARRPLVAAPYIVVDMEN 120
    |||||||
DB 209 nhekkgltkyasspekswtadpysdlekvggridknvsparrplvaaypivvdmen 268
OY 121 ILSKNEOSTONTSETRTISKNTSTRTSTRTSEVHGNAEVANNTSTRTHTSEVHGNAE 180
    |||||||
DB 269 ilsknedsgntdsertliskntststststsevhgnaevnaftddgysvsagfsn 328
OY 181 VHAVALDHSLSAGERTAEWTGMLTADTARLANIRIVNTGTADIVNVLPPTSVLCKN 240
    |||||||
DB 329 sstvaldhslsagertvaetmgltadtarlannirivntgtadivnvlpptsivlgkn 388
OY 241 OTLATITAKENQLSOILAPNNYPSKNAPIALNODDESSPITMANTNOFELEKTKOL 300
    |||||||
DB 389 qlatikakenglsqilapnnypsknapialnaqddstscplmynqfielektqil 448
OY 301 RLDTQVYGNITATVNFENGVRVDTGSNMSEVLPOIETTARIIFNGKDLNVERRIAAY 360
    |||||||
DB 449 rldtqvygniatynfengvrvtsgnmsevlpqlgetarilifngdlnverriaav 508
OY 361 NPSDPLETTKPDMTLKEALKIAFGNEPENGNIQYOGKDIITEFDENFDQOTSNIKNOAE 420
    |||||||
DB 509 npsdplettkpmtlkealkiafgnepngniqyqgkditefdnfdqtsqnknqlae 568
OY 421 LNATNITVLDKIKLNAMKNILIRDKRPHYDRNNTAVGADSEVKEAAREVINSSTEGIL 480
    |||||||
DB 569 lnatnityldkiklnakmmlllrktrhydrnmavgadsevvkeahrevinsstegil 628
OY 481 LNIIDKIRKILSGYIVETEDTEGLEKVINDRYDMLNISLRDQGTFTIDFKKYNKLPPLY 540
    |||||||
DB 629 lniidkirkilsgyiveledteglekevinndrydmlnisslrddgkftidfkynkplply 688
OY 541 ISNPYKYNVAVTKENT 558
    |||||||
DB 689 isnpykynvavtkent 706
    |||||||
```

RESULT 2

AAVS6959 standard; protein: 736 AA.

AAVS6959;

25-APR-2000 (first entry)

B. anthracis MAT-PA protein.

Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA; tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.

OS *Bacillus anthracis*.

XX WO200002522-A2.

XX 20-JAN-2000.

XX 09-JUL-1999; 99WO-US15568.

XX 10-JUL-1998; 98US-0092416.

XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.

XX Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;

XX WPI: 2000-182165/16.

XX N-PSDB; AA256875.

XX Recombinant DNA construct useful as vaccines for anthrax, in producing

XX host cells for analyzing the drugs and agents inhibiting anthrax -

XX Disclosure: Page 34; 35pp; English.

The invention provides a recombinant DNA construct that comprises a vector and at least one nucleic acid (or its fragment) encoding a combination of *Bacillus anthracis* proteins, selected from protective antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA (PA with its secretory signals replaced with those of tissue plasminogen activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine for anthrax and in producing infectious alpha virus particles. These particles, expressing the B. anthracis proteins are useful also as vaccines for anthrax. Host cells transformed with the construct are useful for analyzing the effectiveness of drugs and agents that inhibit anthrax or B. anthracis proteins. The present sequence represents a B. anthracis MAT-PA protein.

XX Sequence 736 AA:

Query Match 96.1%; Score 2758; DB 21; Length 736;
Best Local Similarity 96.8%; Pred. No. 7,4e-188;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

```
OY 1 DNLQLPELKQKSSNRKRSTAGPTVPDRNDGIPDSLEVEGYVVDVKNKRTFLSPWIS 60
    |||||||
DB 150 dnlqlpelkqkssnrkrstagsptvpdrndgipdslevegylvdvknkrtflspwis 209
OY 61 NHEKKGLTKYKSSPEKKSTASDPYSDEPKYGRIDKNVSPARRPLVAAPYIVVDMEN 120
    |||||||
DB 210 nhekkgltkyasspekswtadpysdlekvggridknvsparrplvaaypivvdmen 269
OY 121 ILSKNEOSTONTSETRTISKNTSTRTSTRTSEVHGNAEVANNTSTRTHTSEVHGNAE 180
    |||||||
DB 270 ilsknedsgntdsertliskntststststsevhgnaevnaftddgysvsagfsn 329
OY 181 VHAVALDHSLSAGERTAEWTGMLTADTARLANIRIVNTGTADIVNVLPPTSVLCKN 240
    |||||||
DB 330 sstvaldhslsagertvaetmgltadtarlannirivntgtadivnvlpptsivlgkn 389
OY 241 OTLATITAKENQLSOILAPNNYPSKNAPIALNODDESSPITMANTNOFELEKTKOL 300
    |||||||
DB 390 qlatikakenglsqilapnnypsknapialnaqddstscplmynqfielektqil 449
OY 301 RLDTQVYGNITATVNFENGVRVDTGSNMSEVLPOIETTARIIFNGKDLNVERRIAAY 360
    |||||||
DB 450 rldtqvygniatynfengvrvtsgnmsevlpqlgetarilifngdlnverriaav 509
OY 361 NPSDPLETTKPDMTLKEALKIAFGNEPENGNIQYOGKDIITEFDENFDQOTSNIKNOAE 420
    |||||||
DB 510 npsdplettkpmtlkealkiafgnepngniqyqgkditefdnfdqtsqnknqlae 569
OY 421 LNATNITVLDKIKLNAMKNILIRDKRPHYDRNNTAVGADSEVKEAAREVINSSTEGIL 480
    |||||||
DB 570 lnatnityldkiklnakmmlllrktrhydrnmavgadsevvkeahrevinsstegil 629
    |||||||
```

QY 481 LNIKDRIKILSGYIIEDETEGLKEVINDRYMLNISSLRDQKTFIDFKKYNDKLPly 540
 |||||||
 Db 630 lndkdrlrlslsgyiveideleglkeyindrydmnlsslrqdgktfidtkkndkrlply 689
 QY 541 ISNPNTKVNVAVTKE NT 558
 |||||||
 Db 690 isnpnkvnvyavtkent 707

RESULT 3

ID AAY56960 standard; protein; 763 AA.
 AAY56960;

DT 25-APR-2000 (first entry)
 DE B. anthracis TPA-PA protein.
 DE B. anthracis TPA-PA protein.

KW Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA;
 tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.

OS Bacillus anthracis.

PN WO200002522-A2.

PD 20-JAN-2000.

PF 09-JUL-1999; 99WO-US15568.

PR 10-JUL-1998; 98US-0092416.

PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.

PI Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;

DR WPI: 2000-182165/16.

N-PSDB; AA256876.

PT Recombinant DNA construct useful as vaccines for anthrax, in producing
 PT host cells for analyzing the drugs and agents inhibiting anthrax -
 PS
 XX Disclosure; Page 32; 35pp; English.

CC The invention provides a recombinant DNA construct that comprises a
 CC vector and at least one nucleic acid (or its fragment) encoding a
 CC combination of Bacillus anthracis proteins, selected from protective
 CC antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA
 CC (PA with its secretory signals replaced with those of tissue plasminogen
 CC activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine
 CC for anthrax and in producing infectious alpha virus particles. These
 CC particles, expressing the B. anthracis proteins are useful also as
 CC vaccines for anthrax. Host cells transformed with the construct are
 CC useful for analyzing the effectiveness of drugs and agents that inhibit
 CC anthrax or B. anthracis proteins. The present sequence represents a
 CC B. anthracis TPA-PA protein.

XX Sequence 763 AA;

Query Match 96.1%; Score 2758; DB 21: Length 763;

Best Local Similarity 96.8%; Pred. No. 7.8e-186;

Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 1 DNILQIPELKOKSSNRKRSTSGAPTPVPRDNDGIPDSLEVEGYTVDVKNKRTFLSPWIS 60
 |||||||
 Db 177 dnlqipelkqksnsrkrstsgaptpvprdrndgipdsleveytydvknkrflfispwis 236
 QY 61 NIHEKKGLTKRYSSPEKWSSTADSPYSDFEKKVTGRIDKNVSPEARHPLVAAYPIVHVMEN 120
 |||||||
 Db 237 nlhekkgltkrysspekwsstadspsdfekvtgridknvspearhplvaaypivhvdmen 296

QY 121 IILSKNEDOSTQWNTDSEPTTISKNTSTSRHTSEVHNAEVAHNTSTSRHTSEVHNAE 180
 |||||||
 Db 297 ilsknedgstqntdsestriskntstsrhtsevhnaevahntstsrhtsevhnae 356
 QY 181 VHAVIDHSISLAGERTWAEFTMGINTADTARLANANIRYVNTGAPLYNVLPPTSLVLGN 240
 |||||||
 Db 357 ssvaidhsislagertwaetmgintadtarlananiryvntgaplynvlpptslvlgn 416
 QY 241 QTLATIKAKENQSLQIAPNNYPSKRLAPIALNAODESSTPTMMYNQFLEKTKOL 300
 |||||||
 Db 417 qlatlkakenqslqiapnnypskrlapialnaoddistptlmmynqflekckql 476
 QY 301 RLPTDOVYGNITATYNEPENGVRVDTGSMSEVLPQIETTARIIFNGKDLNVERRIA 360
 |||||||
 Db 477 rldtdgygniatynfengvrvdtsgsmsevlpqiettarilfngkdlnverria 536
 QY 361 NPSDPLETTRKDMTLKALKIAEFNPNGNLOYGKDIRTEPFENFDQTSQNIKNOLAE 420
 |||||||
 Db 537 npsdplettrkdmrtlkalkealfgfnepngnlygqkdltefdnfqgtsqnlknqlae 596
 QY 421 LNAITNTYTVDDTKIKNAKMMILRDKRFHYDRNNIANGADESVYKKAHREYINSGTGL 480
 |||||||
 Db 597 lnaitntylvdtkiknakmllrdrkrfhydrnniavgadesvkkeahrevinsstegl 656
 QY 481 LNIKDRIKILSGYIIEDETEGLKEVINDRYMLNISSLRDQKTFIDFKKYNDKLPly 540
 |||||||
 Db 657 lndkdrlrlslsgyiveideleglkeyindrydmnlsslrqdgktfidtkkndkrlply 716
 QY 541 ISNPNTKVNVAVTKE NT 558
 |||||||
 Db 717 isnpnkvnvyavtkent 734

RESULT 4

ID AAY56958 standard; protein; 764 AA.

XX AAY56958;

DT 25-APR-2000 (first entry)

DE B. anthracis protective antigen (PA) protein.

KW Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA;
 tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.

OS Bacillus anthracis.

PN WO200002522-A2.

PD 20-JAN-2000.

PF 09-JUL-1999; 99WO-US15568.

PR 10-JUL-1998; 98US-0092416.

PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.

PI Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;

DR WPI: 2000-182165/16.

N-PSDB; AA256874.

PT Recombinant DNA construct useful as vaccines for anthrax, in producing
 PT host cells for analyzing the drugs and agents inhibiting anthrax -
 PS
 XX Disclosure; Page 33; 35pp; English.

CC The invention provides a recombinant DNA construct that comprises a
 CC vector and at least one nucleic acid (or its fragment) encoding a
 CC combination of Bacillus anthracis proteins, selected from protective
 CC antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA
 CC (PA with its secretory signals replaced with those of tissue plasminogen

CC activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine
 CC for anthrax and in producing infectious alpha virus particles. These
 CC particles, expressing the B. anthracis proteins are useful also as
 CC vaccines for anthrax. Host cells transformed with the construct are
 CC useful for analyzing the effectiveness of drugs and agents that inhibit
 CC anthrax or B. anthracis proteins. The present sequence represents a
 CC B. anthracis PA protein.

XX
 SQ Sequence 764 AA;

Query Match 96.1%; Score 2758; DB 21; Length 764;
 Best Local Similarity 96.8%; Pred. No. 7.8e-188;
 Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

OY 1 DNLOPELKOKSSNRKRSTAGPTVPDRNDGIPDSLEVEGYVDVKNKRTFLSPWIS 60
 Db 178 dnldpelpkqgssnrkrstagsptvpdrndgipdslevegylvdvknkrtflspwis 237
 OY 61 NIHEKGLTKYKSSPEKSTASDPYSDEKYTGRIKKNVSPEARHPLVAAPYIVHVDMEN 120
 Db 238 nihkkgltkyksspekstasdpysdfevtygridknvspearhplvaaypivhvdmen 297
 OY 121 IILSKNEDQSTQNTDSETRTISKNTSTRTHTSEVHGNAEVAHANTSTRTHTSEVHGNAE 180
 Db 298 ilsknedqscntdsetrtiskntsttrhtsevhgnaevhaasfidigsvsagfsn 357
 OY 181 VHVAIDHSISLAGERTVAETMGINTADTARLNANIRVYNTGTAPIYVAVLPTTSVLGKN 240
 Db 358 ssvaidsislagertvaetmgintadtarlannirvnygtapiyvlpttsvlgkn 417
 OY 241 QTLATIRAKENOLSOILAPNNYPSKNLAPIALNAQDFSSPTITMANYNOGLEKTKQL 300
 Db 418 qtlatikakenglsqilapnnypsknlapialnaqdfssptitlmynqglektql 477
 OY 301 RLDDQVYGNATYNEFNGRVVDGSMWSEVLPQIOETARIIFNGKDLNVERRIAAY 360
 Db 478 rlddqvgynatynefngrvvdgsmwsevlpqioetarilfngkdlnverriaav 537
 OY 361 NPSDLETTKPDMTLEKALITAFGNEPNNGNLOYOGKIDTFEDFNQDOTSQINQOLAE 420
 Db 538 npsdpletktpdmtlekalitafgfnepngnlqygskdtfedfnfdqgtsqnknqlae 597
 OY 421 LNATNITVLDKIKLNAMNLLIRDKRFRHYRNNINAVGADSVYKEAREVINSSTBGL 480
 Db 598 lnatnitylvldkiklnamnlilrdkrfhyrnniavgadesvvykearevinsstegl 657
 OY 481 LNIDKIRKILSGYIVETEDTEGLKEVINDRYDMNLSLRDQGTFTDFKKNKDLPLX 540
 Db 658 lnidkirkilsgyiveitedteglkevindrmdnlsllrqdgkrlidfkknkdlply 717
 OY 541 ISNPNKYVNAVYATKENT 558
 Db 718 isnpnkyvnavyatkent 735

RESULT 5
 AAB47306
 ID AAB47306 standard; Protein: 764 AA.
 XX
 AC AAB47306;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE Wild type B. anthracis protective antigen.
 XX
 KW Lethal factor: LF; immunogen: LF4; protective antigen: PA; DNA vaccine;
 KW humoral; cell-mediated; immune memory response.
 XX
 OS Bacillus anthracis.
 XX
 Key Location/Qualifiers
 FH Peptide 1..29

FT /label= Signal peptide
 FT /note= "Not given in the specification"
 FT Protein 30..764
 FT /label= PA
 FT Peptide 204..764
 FT /label= pcpa

PN W0200145639-A2.

PD 28-JUN-2001.

PE 21-DEC-2000; 2000WO-US34912.

PR 22-DEC-1999; 99US-0171459.

PA (OHIS) UNITV OHIO STATE RES FOUND.

PA (GALL/) GALLOWAY D R.

PI (MATE/) MATECZUN A J.

PI Galloway DR, Mateczun AJ;

DR WPI: 2001-408540/43.

DR N-PSDB: AAC86016.

PT Protecting animal against lethal infection with Bacillus anthracis, by
 PT administering wild type or mutated form of Bacillus anthracis lethal
 PT factor protein or its fragment or a nucleic acid encoding the mutated
 PT protein -

PS Claim 5; Fig 2; 33pp; English.

XX This sequence shows the B. anthracis protective antigen (PA). An
 CC immunogenic fragment of PA, pcpa, can be used to produce an immune
 CC response which protects an animal against lethal infection with
 CC Bacillus anthracis. DNA encoding the B. anthracis PA can be used
 CC in conjunction with DNA encoding the lethal factor (LF) in a
 CC DNA vaccine. Using a DNA vaccine which encodes the mutated LF protein
 CC or fragment alone or in combination with a DNA encoding the PA protein
 CC or its fragment, both components (humoral and cell-mediated) of the
 CC immune system are stimulated, which results in longer term immune
 CC memory response. The combined use of a mutated LF and PA gene or their
 CC fragments results in a higher level of immune response, as judged by
 CC overall serum antibody titers for LF and PA antigens, than the use of
 CC either LF or PA genes in separate immunizations.

XX Sequence 764 AA;

Query Match 96.1%; Score 2758; DB 22; Length 764;
 Best Local Similarity 96.8%; Pred. No. 7.8e-188;
 Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

OY 1 DNLOPELKOKSSNRKRSTAGPTVPDRNDGIPDSLEVEGYVDVKNKRTFLSPWIS 60
 Db 178 dnldpelpkqgssnrkrstagsptvpdrndgipdslevegylvdvknkrtflspwis 237
 OY 61 NIHEKGLTKYKSSPEKSTASDPYSDEKYTGRIKKNVSPEARHPLVAAPYIVHVDMEN 120
 Db 238 nihkkgltkyksspekstasdpysdfevtygridknvspearhplvaaypivhvdmen 297
 OY 121 IILSKNEDQSTQNTDSETRTISKNTSTRTHTSEVHGNAEVAHANTSTRTHTSEVHGNAE 180
 Db 298 ilsknedqscntdsetrtiskntsttrhtsevhgnaevhaasfidigsvsagfsn 357
 OY 181 VHVAIDHSISLAGERTVAETMGINTADTARLNANIRVYNTGTAPIYVAVLPTTSVLGKN 240
 Db 358 ssvaidsislagertvaetmgintadtarlannirvnygtapiyvlpttsvlgkn 417
 OY 241 QTLATIRAKENOLSOILAPNNYPSKNLAPIALNAQDFSSPTITMANYNOGLEKTKQL 300
 Db 418 qtlatikakenglsqilapnnypsknlapialnaqdfssptitlmynqglektql 477
 OY 301 RLDDQVYGNATYNEFNGRVVDGSMWSEVLPQIOETARIIFNGKDLNVERRIAAY 360

|||||
Db rldtdgvygniatytnengrvrvtgtsnwsevlpqiqtetariifngkdliverriaav 537
QY 361 NPSDPLETTRKPDMTLKEALIKAFGEFNEPNCNLOYOGKDIRTEPFNDQOOTSOMIKNOLAE 420
Db 538 npsdplettrkpdmtlkealikaftgtnepnqlgyqgkdtelfndtqgtsqnlknlae 597
QY 421 LNATNIYTVLDKIKLNAKNNILIRDKRFHYDRNNINAVGADESYYKEAREVINSSTEGLL 480
Db 598 lnatniyvtldkiklnaknnlirdkrfhydrnninavgadesvkeahrevinsstegll 657
QY 481 LNIDKDIRKILSGYIYEIEDTSGLKEVINDRIDMLNISSLRDQKTFIDFKKYNKLP 540
Db 658 lndkdrkrlsgyiveiedteglkevindrlydmnlsslrqgkctfidfkynkklply 717
QY 541 ISNPYKVNYYAVTKENT 558
Db 718 isnpykvnnyavtkent 735

RESULT 6

AAR60183 standard: Protein; 903 AA.

AC AAR60183;
DT 04-APR-1995 (first entry)
DE PA(1-725)-----Human CD4 fusion protein coding sequence.
XX
XX Anthrax: Bacillus anthracis; fusion protein; lethal factor;
XX protective antigen; cell killing; targeting; targeting; pathogen;
XX intracellular; HIV; human immunodeficiency virus; toxin;
XX Pseudomonas; exotoxin.
XX
XX Bacillus anthracis.
OS Homo sapiens.
XX
XX WO9418332-A.
PD 18-AUG-1994.
XX
XX 14-FEB-1994; 94WO-US01624.
PF
XX 12-FEB-1993; 93US-0021601.
PR 25-JUN-1993; 93US-0082849.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA
XX Arota N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;
PI
XX WPI: 1994-279753/34.
DR N-PSDB; AAQ70184.
XX
XX Nucleic acid encoding anthrax toxin fusion protein - useful for
PT targeting toxin to specific cells, eg for killing tumour cells
PI or HIV-infected cells
XX
XX Disclosure; Page 100-103; 124pp; English.
XX
XX This sequence is a fusion protein comprising amino acid residues
CC 1-725 of the anthrax protective antigen protein and residues 1-178
CC of human CD4, the portion which binds to gp120 on HIV infected
CC cells. Such fusion proteins may be useful for the specific killing
CC of tumour cells or the killing of cells infected with intracellular
CC pathogens, especially HIV, depending on their components.
XX
XX Sequence 903 AA:

Query Match 96.1%; Score 2758; DB 15; Length 903;
Best Local Similarity 96.8%; Pred. No. 1e-187;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 1 DNQLPELKOKSSNSRKKRSTASGPTVPDRNDGIDPSLEVEGYVDVKNKRTFLSPWIS 60
Db 149 dnqlpelkoksnsrkkrstasgptvpdrndgipdslevegvydvknkrtflspwis 208
QY 61 NHEKKGLTKYKSSPEKWSASPYSDFEYKTVGRIDKNVSEARHPLVAAPYIVHVDMEN 120
Db 209 nhekkgltkykspekwsaspydsdfekvtgridknvspearhplvaapyivhvdmen 268
QY 121 IILSKNEPOSTOMTDESTRISKNTSRRHTSEVGNAAVHANTSRHRTSEVGNAAE 180
Db 269 iilsknedgstomtdestriskntsrrhtsevgnaavhantsrhrtsevgnaae 328
QY 181 VHAVALDHSLSLGEPTWATMGLNTADTARLANIRYVNTGTAPYVNLPTTSVLGKN 240
Db 329 sstvaldhsislsgertwattmglntadtarlanirynvtgtapynvnlptsylvlgn 388
QY 241 QTLATIRAKENQLSQILAPNNYPSKNLADIALNADDESSTPIYMNNOFLELEKTKOL 300
Db 389 qtlatirakenqslsapnnypsknladialnaddesstpiymnnoflelektkql 448
QY 301 RLDTDYVYGNIAIYVNFENGVRVYDTSNMSEVLPQIDETTARIIFNCKDLNIVERRIAAY 360
Db 449 rldtdgvygniatytnengrvrvtgtsnwsevlpqiqtetariifngkdliverriaav 508
QY 361 NPSDPLETTRKPDMTLKEALIKAFGEFNEPNCNLOYOGKDIRTEPFNDQOOTSOMIKNOLAE 420
Db 509 npsdplettrkpdmtlkealikaftgtnepnqlgyqgkdtelfndtqgtsqnlknlae 568
QY 421 LNATNIYTVLDKIKLNAKNNILIRDKRFHYDRNNINAVGADESYYKEAREVINSSTEGLL 480
Db 569 lnatniyvtldkiklnaknnlirdkrfhydrnninavgadesvkeahrevinsstegll 628
QY 481 LNIDKDIRKILSGYIYEIEDTSGLKEVINDRIDMLNISSLRDQKTFIDFKKYNKLP 540
Db 629 lndkdrkrlsgyiveiedteglkevindrlydmnlsslrqgkctfidfkynkklply 688
QY 541 ISNPYKVNYYAVTKENT 558
Db 689 isnpykvnnyavtkent 706

RESULT 7

AAV56961 standard: protein; 569 AA.

AC AAV56961;
DT 25-APR-2000 (first entry)
DE B. anthracis PA63 protein.
XX
XX Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA;
XX tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.
XX
XX Bacillus anthracis.
OS
XX WO200002522-A2.
PN 20-JAN-2000.
XX
XX 09-JUL-1999; 99WO-US15568.
PF 10-JUL-1998; 98US-0092416.
PR
XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
PA
XX Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;
PI
XX WPI: 2000-182165/16.
DR N-PSDB; AA256877.
XX
XX Recombinant DNA construct useful as vaccines for anthrax, in producing

PT		host cells for analyzing the drugs and agents inhibiting anthrax -
PX		
PS		Disclosure: Page 35; 35pp; English.
XX		
CC		The invention provides a recombinant DNA construct that comprises a
CC		vector and at least one nucleic acid (or its fragment) encoding a
CC		combination of Bacillus anthracis proteins, selected from protective
CC		antigen (Pa), Mar-Pa (Pa with its secretory signals removed), rPA-Pa
CC		(PA with its secretory signals replaced with those of tissue plasminogen
CC		activator) and Pa63 (63 kDa PA). The construct is useful as a DNA vaccine
CC		for anthrax and in producing infectious alpha virus particles. These
CC		particles, expressing the B. anthracis proteins are useful also as
CC		vaccines for anthrax. Host cells transformed with the construct are
CC		useful for analyzing the effectiveness of drugs and agents that inhibit
CC		anthrax or B. anthracis proteins. The present sequence represents a
CC		B. anthracis Pa63 protein.
XX		
SQ	Sequence	569 AA:
	Query Match	92.8%; Score 2664; DB 21; Length 569;
	Best Local Similarity	96.7%; Pred. No. 2,5e-181;
	Matches	521; Conservative 2; Mismatches 16; Indels 0; Gaps 0.
OY	20	STSGAPVPDRDNDGIPDSLEVEGYTVDYKNRRTLSPMISINHEKKGLTKXSPEKWS 79
DB	2	sfsagptvprdrndgipdsleveytvdyknkrftfslmsnhbkgytkykspskws 61
OY	80	TASDPYSDEKYTGRIKDKNVSPEARHPVAAAPRVHVDENIIISKNEQSQNDSER 139
DB	62	tasdpysdekytgrikdknvspearhpvaaprvlvdmennllsknedsqstntser 121
OY	140	TISKMTSTRTHTSSEVHGNAEVAHNANTSTRTHTSSEVHGNAEVAHAVIDSLAGERTTA 199
DB	122	tiskmtstrthtssevhgnaevhaafidfgsvsaegfsmnsnstvaldmslagertta 181
OY	200	ETMGINTADTARLANANIRVVTGARIVWLPTSLVLGSKNOTLATIKAKENQLSQTILAP 259
DB	182	etmgintadtarlananirvyvtgarivwlptsylvlgsknotlatikakenglsqtlap 241
OY	260	NNYYSKNLAPIALNAODEFSRPLTMANNQGLELEKTQKLMDPDQVYGNTATYFENG 319
DB	242	nnyysknlapialnaodefsrpltmannoglleektqklmdpdqvvgntatynfeng 301
OY	320	RVRVDTGSNMSSEVLRPOIOTFTARIIFNGKDLNVERRIAAVNSPDLRTKRPDMTLKEAL 379
DB	302	rvtvdgtsgnmsevlpqiagetarilfnngkdlnlverriaavnsdpdlrtkpdmtlkeal 361
OY	380	KTAGFNEBPNGLAYOGKADTFEDFNFDQOSTQNIKNOLAELNATNYITVLDRIKLNAM 439
DB	362	ktagfnebpnglayyqgdktfedfnfdqbstqnknqlaelnatniylvdriklnamk 421
OY	440	NLLIDDKRRHYRNNNIYAGADESVYKKEAREYINSSTBELLIANDKDIKLSGITVEIE 499
DB	422	nlllddkrfhydnmlayagadesvykkearelynsssteglllnlkdkrlklsiyiveie 481
OY	500	DTEGLEYVINDRYDMLNISLSRODGKTFIDFKKYNDKLPLYTISNPYKNVYAVERENT 558
DB	482	dteglekeyvindrydmnlslsrtdgkttfidfkkyndkrlplytispnykvnyavkent 540
RESULT	8	
ID	AAR60193	
XX	AAR60193 standard; Protein; 719 AA.	
AC	AAR60193;	
XX		
DT	04-APR-1995 (first entry)	
XX		
DE	Modified protective antigen of Bacillus anthracis.	
XX		
XX	Anthrax; Bacillus anthracis; fusion protein; lethal factor;	
XX	protective antigen; cell killing; targeting; targeting; pathogen;	
XX		

```

KW intracellular; HIV; human immunodeficiency virus; toxin.
XX
OS Bacillus anthracis.
XX
PN MO94I8332-A.
XX
PD 18-AUG-1994.
XX
PF 14-FEB-1994; 94MO-US01624.
XX
PR 12-FEB-1993; 93US-0021601.
PR 25-JUN-1993; 93US-0082849.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Arora N, Kimpel K, Leppla SH, Nichols PJ, Singh Y;
XX
DR WPI; 1994-279753/34.
XX
DR N-PSDB; AAQ70189.
XX
PT Nucleic acid encoding anthrax toxin fusion protein - useful for
PT targeting toxin to specific cells, eg for killing tumour cells
PT or HIV-infected cells
XX
PS Example 6; Page 114-115; 124pp; English.
XX
CC The sequence encoding the protective antigen of Bacillus anthracis
CC may be used in the construction of a nucleic acid which encodes a
CC fusion protein comprising the anthrax protective antigen binding
CC domain of the native anthrax lethal factor and a sequence encoding an
CC activity inducing domain of a second protein. Such fusion proteins
CC are useful for the specific killing of tumour cells or the killing of
CC cells infected with intracellular pathogens, especially HIV,
CC depending on the second component. The protective antigen and other
CC toxins require proteolytic cleavage to acquire activity. Since some
CC cells infected with an intracellular pathogen possess an active
CC protease with quite a narrow substrate specificity e.g. HIV, the
CC protease cleavage site found in the native toxin is replaced with an
CC intracellular pathogen specific protease site (See AR60184-88). The
CC protease in cells that are infected with an intracellular pathogen
CC cleaves the modified toxin which is then rendered active and kills
CC the cell. This sequence is a modified Bacillus anthracis protective
CC antigen which has the amino acids originally at positions 162-171
CC replaced with the HIV protease cleavable sequence described in
CC AR60186.
XX
SQ Sequence 719 AA;
XX
XX
Query Match 89.9%; Score 2581; DB 15; Length 719;
Best Local Similarity 90.9%; Pred. No.2.8e-175;
Matches 511; Conservative 6; Mismatches 21; Indels 24; Gaps 2;
1 DNLDPELKKQKSSN---RKKRSTAGCTVVDNRDNGDIPDSLEVEGYVVDKNNRPTLS 56
|||||:::|
149 dnlpelkqksentatlmmgngfipgtprcdndgipdsleveyvvdcknkrftls 208
57 PWISINHEKKGITKRSSEKMWSTASDPISDEKYTGRIKDKVSPDARHPLVAAPYVHV 116
PWISINHEKKGITKRSSEKMWSTASDPISDEKYTGRIKDKVSPDARHPLVAAPYVHV 116
209 PWISINHEKKGITKRSSEKMWSTASDPISDEKYTGRIKDKVSPDARHPLVAAPYVHV 268
117 DNEITILSNEQOSQONDSERTTSKNTSTRTTSEVHGAAEVHANSTSRHTSEVH 176
dneitilsknedgsqndsertrtskntstsrhtsevhgnaevhasfdlgsvsagf 328
269 dneitilsknedgsqndsertrtskntstsrhtsevhgnaevhasfdlgsvsagf 328
177 GNAEVHAAVAIDHSLSLAGERTWAEFTMGINTADTARLANIRRVNNGTAPIYVNVLTPTSTLV 236
:|||||
329 snnsstvaldshslagertwaeftmgintadtarlanirrvnngtapiyvnvltptstlv 388
237 LGRNQTLATIRKKEQOLSIILAPNNYVPSKNIAPIALNADDFSSPTPTTMANYNOFLELEK 296
lgrnqtlatirkkeqolsiilapnnvyvpskniapialnadddfssptpttmannynoflelek 296
389 lgrnqtlatirkkeqolsiilapnnvyvpskniapialnadddfssptpttmannynoflelek 439

```

Sequence	884 AA:
Query Match	20.7%; Score 593; DB 17; Length 884;
Best Local Similarity	26.4%; Pred. No. 1e-33;
Matches 194; Conservative	95; Mismatches 213; Indels 232; Gaps 25;
1	DNLQLEPELKQKSS-----NSRRKRSTASGPTVPDRDNDGIPDSLEVEGYVD 47
Db	delrnpelkkesgeflakpksklnftfkkmkreided--tdtdgdsipdlweengytl- 244
QY	48 VKNKRTFLSPVTSNIHHEKGLTKRYSSPEKWSADPDSDFEKYGRIDKNVSPARHPL 107
Db	245 ---grrlrvkwdsd-askgyklkvsnpleshtvdydpyldyekaardldlnaketfnpl 300
QY	108 VAAVYIVHVDMENILSKNKEDOSTONTDSEPTIKSTNSTSRHTSEVHGAIEVANTST 167
Db	301 vaafsvsvsmekvllspneis-----nsvsesstchw 334
QY	168 SRHTSEVHGAIEVH-----AVADHSLSLAGERTWAETG---LNTADTRALNAN 215
Db	335 sylvntegasseveagipgkqisfgsvngvnielvae--wgtstgntsgfntasagyl 392
QY	216 IRYVVTGAPRIYVNVPTPSVLGKQOTLATIKAKENQSLIAPNNVYPSKALAPALNA 275
Db	393 vrymvgcgaalydvkptsfvyl-ndtlatlataknsstalnispesypkkqggniafts 451
QY	276 ODDFSSTPTMNVNOFLELEKTQOLRLPDQVYGNIATYVFNENGVRVDTGSGMSEVLPO 335
Db	452 mddfnshpclkknkqvadlllnkpmmltelngdg---vykldctngnlvtgsewngvlyq 508
QY	336 IOETTARIIFNGKDLNIVERRIAAVNPSDPLETTRKPDMTLKEALKIAF--GFNEPENGMLQ 393
Db	509 lkaktesallvddge-rvaeekrvaaakdyenpedkt-psltlkdaalklvsydpdelksetegilly 566
QY	394 YQGKITEPDE--NPDQOTSQNIKQOLAEL-----NATNIYIVLDKIKLNAKANLILRDK 446
Db	567 yknkpllyessvmcyldentakevtkqjndctgkfkdvshlydv---kltpkmvntlk-1 621
QY	447 RHHYRNNINAVGADSVYVEKAHREYINSTEKG-----LLN----- 482
Db	622 silyln---aesndnsigkvtntnlvsgngnkkqysnmpdanltnlntdaqeklnknd 678
QY	483 ----- 482
Db	679 yylisymksektgcetlidgeilyptlctkvnvknknyrlidiahnkspisslhkt 738
QY	483 -----IDKDIRKLKLSGYIEIETEGE-----KEYI 508
Db	739 ndelrlfwddisltvdasakpenltdseikqlyrsyglkled--qllldkkgjnygeftl 796
QY	509 ND-----RY-----DMLNLTSSLRDCKTGFIDEFKY--NDKRLPLY 540
Db	797 neasfnieplqnytkylevtysseljgnvnsdclsesdklykdgctlkxfdfckysknglyft 856
QY	541 ISNPNVKVVVYAVT 554
Db	857 dsglnwdfnfnait 870
RESULT 10	
AAM60224	
ID AAM60224	standard; Protein; 880 AA.
AC AAM60224;	
XX	
DT 28-SEP-1998	(first entry)
XX	
DE Bacillus thuringiensis insecticidal toxin 177C8.	
XX	
KW Insecticide; pesticide; toxin; delta-endotoxin;	
KM biological control; lepidopteran; coleopteran.	
XX	

OS	Bacillus thuringiensis strain PS177C8 (NRRL B-21867).
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 253
FT	Misc-difference 675 /note= "encoded by YTA"
FT	Misc-difference 846 /note= "encoded by AC"
XX	Misc-difference 846 /note= "encoded by RAA"
XX	
PN	MO9818932-A2.
XX	
PD	07-MAY-1998.
XX	
PF	30-OCT-1997; 97WO-US19804.
XX	
PR	30-OCT-1996; 96US-0029848.
XX	
PA	(MTCO) MTCOGEN CORP.
XX	
PI	Dullum CJ, Feltelson JS, Loewer D, Muller-Cohn J;
PI	Narva KE, Schmelts JL, Schnepf HE, Schwab G, Stamp L;
PI	Stochhoff BA;
XX	
DR	WPI: 1998-272226/24.
XX	N-PSDB: AAV30307.
XX	
PT	Bacillus thuringiensis isolates - used for producing pesticidal
PT	toxins and nucleotide sequences for control of lepidopterans and
PT	coleopterans
XX	
PS	Claim 5; Page 81-84; 139pp; English.
XX	
CC	This polypeptide comprises a novel soluble toxin of Bacillus
CC	thuringiensis (B.t.) strain PS177C8 (NRRL B-21867). The toxin
CC	belongs to a novel family of B.t. toxins that have toxicity
CC	against non-mammalian pests. Its amino acid sequence was deduced
CC	from a novel DNA fragment (see AAV30307) obtained by PCR from
CC	cellular genomic DNA of PS177C8. Disclosed and claimed are novel
CC	B.t. isolates and toxins (see AAW60218-32) that have activity against
CC	lepidopteran and/or coleopteran pests, isolated genes, probes
CC	and primers (see AAV30288-321 and AAT99734-87) useful for production
CC	of the toxins and for the identification and characterisation of
CC	these toxins, and transformed hosts, particularly plant and
CC	bacterial hosts. The invention provides 8 entirely new families of
CC	toxins from B.t. isolates. The toxins have the additional ability
CC	to form pores in cell membranes, and can be used to facilitate
CC	entry of a second agent into a target cell.
XX	
XX	Sequence 880 AA:
SO	
	Query Match 20.6%; Score 591.5; DB 19; Length 880;
	Best Local Similarity 29.9%; Pred. NO. 1.3e-33;
	Matches 180; Conservative 96; Mismatches 215; Indels 107; Gaps 22.
QY	1 DNLQLPELKOKSS-----NSRRKRSTSGAPTPPDRNDGIPDSLEVEGYTD 47
Db	186 delrlnpefnkkesqgfakpkslnlftqkmkreided---tltdgdsipdlweengytl- 241
QY	48 VKNKRTFLSPWISNIHEKKGLTKYVSSPEKWSSTASDPYSDFEKVTGRIDKNVSPARHPL 107
Db	242 ---qnrflavkwddsf--aaskgykfkfvsnpleshtvgdpytdykaaardldlsmakctfnpl 297
QY	108 VAAVYIVVHVDENILLSNEDOSTONTOSEFTTISKNTSTKTRHSEVGNNAEVNANST 167
Db	298 vaafpsvuvsmekvllspnrls-----nsveshsstlw 331
QY	168 SRTHTSEVGNNAEVH-----AAVAIDHSLSLAGEPTWAETMG-----LNTADTARLIAN 215
Db	332 sytlntegssvaeqlypdklsgfsvsvnyghsetvaeq--wgtsctgntsgfntlsasayllhan 389
QY	216 IRYVNTGTAPILYVNLPTTSLVGLGNQTLATIRAKENQSLQILAPNNYVPSKMLAETALNA 275

Dd 390 vrynmvgalgydvkptcsfvl-mndlatlctaksnsalnispgeyspkqngjaits 448
Qy 276 QDDSSPTITMNYNOFLELEKTKQLRLDTPDOVYGNATYNEENGSRVYDGSMSEVLPO 335
Db 449 mddfnshpctlnkkqvdlnlmkpmmljetnqtgd---vykikdchnlvitgswngvlyiq 505
Qy 336 IOETTRARIIFNGKDLNLVERRIAVNPSDPLETTKKPDMTLKEALKIAF--GFNPNGNLQ 393
Db 506 ikakasslivddoge-rvaekirvaadvenpeckt-pslltkdaIkisypdelkeieglly 563
Qy 394 YOGKDITEEDF--NFDDQTSOINIKQNALEL-----MATNIITYVDLKIKLNAKNILIRDK 446
Db 564 yknpiysswmtlyidentakevkctqindttgkfkdvshtyd----klcpkmvtlk-1 618
Qy 447 RHHYRNRIIANGADESVYKEAHREVINSTG-----LLNTID-----KDIRK 489
Db 619 sllydn---aesndsisgkwcthlivsggnnkqkyasnmpdanltntdaqeklnry 675
Qy 490 ILSGIVEIEDTEGLKEYINDRYDMLANTS-SLRDDGTFIDFKRKYNDKLPYISPNYKV 548
Db 676 yslymksekhntgcetldgelipctektvnvndnykrldishnhik-----snplssi 730
Qy 549 NV 550
Db 731 hl 732

RESULT 11
AAVS92277
ID AAVS92277 standard; Protein; 881 AA.
XX
AC AAVS92277;
XX
DT 18-APR-2000 (first entry)
DE MIS toxin from B. thuringiensis strain PS177C8.
XX
KW Bacillus thuringiensis; toxin; endotoxin; pesticidae; plant pest;
KW Lepidoptera; cleopterans.
XX
OS Bacillus thuringiensis.
XX
FH Key Location/Qualifiers
FT Misc-difference 846 /note= "unknown"
XX
PN MO9957282-A2.
XX PD 11-NOV-1999.
XX PF 06-MAY-1999; 99WO-USO9997.
XX PR 06-MAY-1998; 98US-0073898.
PA (MYCO) MYCOGEN CORP.
XX Feitelson JS, Schepf HE, Narwa KE, Stockhoff BA, Schmeltz J;
PI Lower D, Dullum CJ, Muller-Cohn J, Stamp L, Morrill G;
PI Finstad-Lee S;
DR WPI: 2000-096811/08.
DR N-PSDB: AAZ58786.

New polynucleotides encoding pesticidally active proteins, useful for transforming plants for controlling pests -
Disclosure: Page 68-71: 104pp: English.

CC are useful for transforming plants for controlling plant pests; for
 CC designing primers and probes useful for the identification and
 CC characterization of genes which encode pesticidal toxins. The present
 CC sequence represents a B.t. toxin.

XX Sequence 881 AA:

Query Match 20.6%; Score 591; DB 21; Length 881;
 Best Local Similarity 29.9%; Pred. No. 1.4e-33;
 Matches 180; Conservative 97; Mismatches 218; Indels 108; Gaps 22;

```

OY 1 DNQLPELKOKS-----NSRKRSTSGPTVPDRDNDGIPDSEVEGYTD 47
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 186 delrmpfinkkeqgefakpsklnlftqmkreided---tdtdgdsipdlweengytl- 241
OY 48 VKNKRFLSPWISNIEHKKGLTKYKSSPEKSTASDPYSDPEKVTGRIDKNVSPRARHPL 107
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 242 ---gnrlavkwddsl-askgytkfvsnpleshvqdpdydekaardldlsnaketfnpl 297
OY 108 VAAVPIVHVMENIILSKNEDOSTONTDSETRISKNTSRTHTSEVHGNAEVAHNTST 167
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 298 vaafpsvnmekvllspnenls-----nveshsstlw 331
OY 168 SRTHTSEVHGNAEVAH-----AVADHSLSLAGERTMAETMG-----LNTADTARLAN 215
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 332 syntlegasveaglgpkjglsfgvsvngqshetvage--wgtstgntsgfntasaagylnan 389
OY 216 IRYVNTGTAPIYVNLPTTSLVLGKNQTLATIKAKENQSOILAPNNVYPSKNLAPYALNA 275
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 390 vrynnvgtgalydvkptstfvl-nddtlatlctaknsstalnlspsgesyppkkgnglaits 448
OY 276 QDDFSSTPTMTMNOLELEKTKQLRLDQYGNATATNFENGRRVVDGSMWSEVLPQ 335
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 449 mdfnfshpiltlnkkqvdnlmnpmmletnqtdg---vykikdthgnlytvgewngvlyq 505
OY 336 IOETTRAIIFNGKDLNVERRIAANVPSDPLETTKPDMLKEALIAF--GFNEPENGQLQ 393
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 506 lkaktasliivdage-rvaekrvaakdyenpedkt-psltlkdaiklsypdelieiglyly 553
OY 394 YGKQDITEFDF--NFDQOTSQNIKNQLAEL-----NATNIYTVLDKIKLAKANNILIRDK 446
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 564 ynkprlyessvmtlydentakevtkqindttgkfkdvshlydv---kltpkmvltk-1 618
OY 447 RHHYDRNNIIVAGADESVVKAHREVINSSTEG-----LLNLD-----KDIR 488
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 619 silydn---aesndnsigkwtntnlivsgngngkqysnmpdanlclntdaqeklnkrd 675
OY 489 KILSGYVEIEDTEGLKEVINDRYDMLNIS-SLRDQKTFIDFKKYNDKLPVLSNPYK 547
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 676 yylslymksekntqceitldgeilyptlctkvnnkdnkykrldliahnlk-----snplss 730
OY 548 VNV 550
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 731 lnl 733

```

RESULT 12
 AAM19509
 ID AAM19509 standard; Protein; 884 AA.

XX AAM19509;
 XX 15-APR-1998 (first entry)
 XX B. cereus VIP1A(a) protein sequence.
 DE Vegetative insecticidal protein: Bacillus cereus strain AB78; plant;
 XX Insect; Sesamia nonagrioides; maize; corn borer; toxin.
 XX Bacillus cereus.
 OS WO9726339-A1.
 XX PN

XX 24-JUL-1997.
 PD 23-DEC-1996; 96WO-EP05828.
 XX 15-JAN-1996; 96GB-0000786.
 XX (NOVS) NOVARTIS AG.
 PA Gay PB;
 PI WPI: 1997-385342/35.
 DR N-PSDB: AAT73994.
 XX
 PT Protecting plants against insects of the genus Sesamia using
 PT Bacillus toxic proteins - applied directly or expressed as
 PT heterologous protein by the plant, also transgenic plants expressing
 PT both Cry and Vip type toxins
 PS Claim 8; Page 35-39; 168pp; English.

CC This is the amino acid sequence of the 100 kD vegetative insecticidal
 CC protein (VIP) 1A(a) from Bacillus cereus strain AB78. The protein can
 CC be used in a new method for protecting plants, and their progeny, against
 CC insects of the genus Sesamia by direct or indirect application to the
 CC plant (or seed or growing area). The protein is especially useful to
 CC protect maize plants against the Mediterranean corn borer
 CC (S. nonagrioides).

XX Sequence 884 AA:

Query Match 20.5%; Score 589; DB 18; Length 884;
 Best Local Similarity 30.2%; Pred. No. 2e-33;
 Matches 180; Conservative 94; Mismatches 215; Indels 108; Gaps 22;

```

OY 1 DNQLPELKOKS-----NSRKRSTSGPTVPDRDNDGIPDSEVEGYTD 47
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 189 delrmpfinkkeqgefakpsklnlftqmkreided---tdtdgdsipdlweengytl- 244
OY 48 VKNKRFLSPWISNIEHKKGLTKYKSSPEKSTASDPYSDPEKVTGRIDKNVSPRARHPL 107
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 245 ---gnrlavkwddsl-askgytkfvsnpleshvqdpdydekaardldlsnaketfnpl 300
OY 108 VAAVPIVHVMENIILSKNEDOSTONTDSETRISKNTSRTHTSEVHGNAEVAHNTST 167
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 301 vaafpsvnmekvllspnenls-----nveshsstlw 334
OY 168 SRTHTSEVHGNAEVAH-----AVADHSLSLAGERTMAETMG-----LNTADTARLAN 215
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 335 syntlegasveaglgpkjglsfgvsvngqshetvage--wgtstgntsgfntasaagylnan 392
OY 216 IRYVNTGTAPIYVNLPTTSLVLGKNQTLATIKAKENQSOILAPNNVYPSKNLAPYALNA 275
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 393 vrynnvgtgalydvkptstfvl-nddtlatlctaknsstalnlspsgesyppkkgnglaits 451
OY 276 QDDFSSTPTMTMNOLELEKTKQLRLDQYGNATATNFENGRRVVDGSMWSEVLPQ 335
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 452 mdfnfshpiltlnkkqvdnlmnpmmletnqtdg---vykikdthgnlytvgewngvlyq 508
OY 336 IOETTRAIIFNGKDLNVERRIAANVPSDPLETTKPDMLKEALIAF--GFNEPENGQLQ 393
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 509 lkaktasliivdage-rvaekrvaakdyenpedkt-psltlkdaiklsypdelieiglyly 566
OY 394 YGKQDITEFDF--NFDQOTSQNIKNQLAEL-----NATNIYTVLDKIKLAKANNILIRDK 446
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 567 ynkprlyessvmtlydentakevtkqindttgkfkdvshlydv---kltpkmvltk-1 621
OY 447 RHHYDRNNIIVAGADESVVKAHREVINSSTEG-----LLNLD-----KDIR 488
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 622 silydn---aesndnsigkwtntnlivsgngngkqysnmpdanlclntdaqeklnkrd 678
OY 489 KILSGYVEIEDTEGLKEVINDRYDMLNIS-SLRDQKTFIDFKKYNDKLPVLSNP 544

```

Db 679 yylslymkseknktgeitldgeiypitktvvnkdkykrldliahnik-----snr 730

RESULT 13
AA046712
ID A046712 standard; Protein: 884 AA.

XX
AC A046712;
XX
DT 02-JUN-1998 (first entry)
XX
DE 100 kDa VIP1A(a) toxin of *Bacillus cereus*.
XX
KW Vegetative insecticidal protein; VIP; expression; maize; protection;
KW plant; *Ostrinia furnacalis*; Asian Corn Borer; Cry toxin; VIP toxin;
KW recombinant; *Bacillus thuringiensis*; transgenic plant; resistance;
KW insect attack; *Sesamia*; maize; cereal crop.
XX
OS *Bacillus cereus*.
XX
PN M09746105-A1.
XX
PD 11-DEC-1997.
XX
PE 27-MAY-1997; 97WO-EP02737.
XX
PR 06-JUN-1996; 96GB-0011777.
XX
PA (NOVS) NOVARTIS AG.
XX
PI Hunter B, Suwantaradon K, Utdewilligen WPM;
XX
DR WPI: 1998-041787/04.
DR N-PSDB; AAV16166.
XX
PT Administration of *Bacillus* sp. toxin protein, especially Cry or
PT vegetative insecticidal protein (VIP) protein to plants - useful for
PT protection against attack by Asian Corn Borer (*Ostrinia furnacalis*)
XX
PS Claim 8; Pages 43-46; 175pp; English.
XX
CC The present sequence represents a 100 kDa vegetative insecticidal
CC protein 1A(a) (VIP1A(a)), and is derived from *Bacillus cereus* strain
CC AB78. The protein is used in a method for protecting plants and their
CC progeny against damage caused by *Ostrinia furnacalis* (Asian Corn Borer).
CC The protein is directly or indirectly applied to the plant, plant seed
CC or growing area of the plant. Cry toxins can also be used in the same
CC way, in place of VIP toxins. The Cry or VIP toxins and genes are used,
CC especially inside recombinant *B. cereus* or *B. thuringiensis* strains,
CC to produce plants protected against Asian Borer pests. Transgenic plants
CC protected against Asian Corn Borer can be used to produce seed and
CC progeny also resistant to insect attack. Plants expressing both a
CC Cry-type and a VIP toxin gene can also protect against *Sesamia* pests.
CC The method and compositions are especially used for protecting maize but
CC may also be used to protect other cereal crops against Asian Corn Borer
CC attack.
XX
SQ Sequence 884 AA;

Query Match 20.5%; Score 589; DB 19; Length 884;
Best Local Similarity 30.2%; Pred. No. 2e-33;
Matches 180; Conservative 94; Mismatches 215; Indels 108; Gaps 22;

QY 1 DNLQPELKKOKS-----NSRRKRSAGPTVPDRDNDGIPDSLEVGTYVD 47
Db 189 delrnpelnkkesgeflkpsklnlftqkmkreidd-----ldtdgdsipdlweengylt- 244
QY 48 VKNKRTFLSPWISNIHEKGLTKYKSSPEKSWSTASDPYSDEKVTGRIDKNVSPEARHPL 107
Db 245 ---qurlavkwddsl-askgytkfkfvsnpleshtvgdpytdyekaardldlslakctfnpl 300

QY 108 VAAPIVHVDMENILSKNEOSTONTSETRTTSKNTSTSKRTHTSEVHGNAEVHANTST 167
Db 301 vaafpsvvnsmekvllspnshls-----AAVAIDHSLAGERTAEFMG-----LNTADARLRLVAN 215
QY 168 SRTHTSEVHGNAEVH-----AAVAIDHSLAGERTAEFMG-----LNTADARLRLVAN 215
Db 335 sylvtegasveagigpkgsifgvsynghsetvage--wgsstgntsgfntasagylhan 392
QY 216 IRYVNTGTAPITVNLPTTSLVYGKNGQTLATIKAKENOLSQLAPNNYPSKNLAPIALNA 275
Db 393 vrynnvgtgalydvkptsfv1-mndlatltakenscalnlspegesypkkgqgjafts 451
QY 276 QDDSSFTITMANYNLFLEKTKQLRLDTDOYGNIAITNEENGVRVDTSSNMSEVLPO 335
Db 452 mddfnshpdltnkqvdnlhnmkpmletnqtdg---vkkldtghnltvgewngvlyq 508
QY 336 IOETARIIFNGKDLNVERRIAANVPSPDETTPPDWTLKALKIAF--GFNEPENGMLQ 393
Db 509 lkaktasllivddge-rvekrtvaakdyenpeokt-psltlkdxalklisydeltkeleglly 566
QY 394 YQKNDITEFDF--NEDOOTSQNIKNQLAEL-----NATNITVLDKIKLNKMNILARDK 446
Db 567 yknkpiyessvmtlyldentakevtqldntgtgkfkdvshlydv---kltpkmvltk-1 621
QY 447 RFHYDRNNIAYGADESIVYKEAHREYINSTEG-----LLNID-----KDIR 488
Db 622 sllydn---aesndnslygtcnlnlvsngmgkqysnnpdanlntldageklxnkr 678
QY 489 KILSGYIEIEDTEGLKEVINDRYDMLNIS--SLRODGKTFDFKYNKDLPLYSNP 544
Db 679 yylslymkseknktgeitldgeiypitktvvnkdkykrldliahnik-----snr 730

RESULT 14
ID AAR91245
AA091245
ID AAR91245 standard; Protein: 1346 AA.

XX
AC AAR91245;
XX
DT 14-AUG-1996 (first entry)
XX
DE VIP2A(a) and VIP1A(a) fusion protein.
XX
KW Pesticide; insecticide; biological control agent; Lepidoptera;
KW Coleoptera; transgenic plant; maize; insect resistance;
KW western corn rootworm; *Dibrotica virgifera* var. *virgifera*; VIP.
XX
OS *Bacillus cereus* strain AB89 (NRRL B-21058).
XX
PN M09610083-A1.
XX
PD 04-APR-1996.
XX
PE 27-SEP-1995; 95WO-EP03826.
XX
PR 05-JUN-1995; 95US-0463483.
PR 28-SEP-1994; 94US-0314594.
XX
PA (CIBA) CIBA GEIGY AG.
XX
PI Carr B, Desai NM, Duck NB, Estruch JF, Kostichka K;
PI Koziel WG, Mullinsma, Nye CJ, Warren GW;
XX
DR WPI: 1996-200921/20.
DR N-PSDB; AAT13944.
XX
PT *Bacillus* strain producing insecticidal protein during vegetative
PT growth - used in the control of Lepidoptera and Coleoptera pests
XX
SQ Claim 30; Page 156-61.; 242pp; English.
XX
CC A fusion protein (AAR91245) is composed of the VIP2A(a) auxiliary
CC protein (AAR91238) and insect-specific VIP1A(a) protein (AAR91239) of

CC Bacillus cereus strain AB38. It is the product of a fusion gene
 CC (AA113944) constructed from the vip coding sequences. The fusion
 CC protein can be expressed e.g. in bacterial cells for use as
 CC biological control agents having improved activity or target range,
 CC or in transgenic plants, esp. maize, to increase insect resistance.
 CC VIPA(a) shows activity against western corn rootworm.

XX Sequence 1346 AA;

Query Match 20.5%; Score 589; DB 17; Length 1346;

Best Local Similarity 30.2%; Pred. No. 3.7e-33;
 Matches 180; Conservative 94; Mismatches 215; Indels 108; Gaps 22;

OY 1 DNLOPELKOKS-----NSRKRSTSGPTVPDRDNGIPDSLEVEGYTD 47
 DB 651 delrnpelfkkesqeflakpsklnlftqkmkreided---tdtdgsdipdlweengytl- 706
 OY 48 VKNKRRTFLSPWISNIEHKGLTKYKSSPEKWSSTASDPYSDFEKTGRIDKNVSPEARHPL 107
 DB 707 ---gnrlavkwddsl-askgytkfvsnpleshvtdydytdekaardldlsnaketfnpl 762
 OY 108 VAAVPIVHVMENIILSKNEDOSTONTDSETRTISKNTSTSRHTSEVHGNAAEVHANTST 167
 DB 763 vaafpsvnmekvllspnenls-----nsveshsstnw 796
 OY 168 SRHTSEVHGNAAEVH-----AVAIDHSLAGERTMAETMG-----LNTADTARLNNAN 215
 DB 797 syntlegasveaglgprkgsifgvsvnyqhselvaqe--wgtsqtsqfntlaasaylnan 854
 OY 216 IRVNTGTAPRYNVLPRTSLVGLKNOTLATIKAKENOLSQILAPNNVYPSKNLAPILNA 275
 DB 855 vrynnvgtalgaydvkptstfvl-ndtiatlatsnstaalinspsesyprkkgngjaits 913
 OY 276 QDDFSTPTITMANYNOFLLEKTKQLRLDQVYGNITATYFENGRRVAVDTGSNMSEVLPQ 335
 DB 914 mddfnspitltnkqvdnltnkmmletnqtdg---vykikdthgnivtggengylyq 970
 OY 336 IOETTRARIIFNGKDLNVERRIAANVPSDPLETTKPDMTKEALKIAF--GFENPGNQLQ 393
 DB 971 ikaktaslivddge-rvaekrvaakdyenpedkt-psltlkdaiklsypdeikeleglly 1028
 OY 394 YOGKDITEFDF--NFDQOTSQINIKNOIAEL-----NATNIYTVLDKIKLAKNMILIRDK 446
 DB 1029 yknkprlyessvmytidentakevtkqindtqkfkdvshlydv---klrpkmmvltk-1 1083
 OY 447 RFHYDRNNIIVAGADESVVKAHREVINSSTEG-----LLINID-----KDIR 488
 DB 1084 silydn---aesndnsigkwtntnivsgnngkkyssnmpdanlntdaqeiklnkrd 1140
 OY 489 KILSGYIVEIEDTEGLEKEVINDRYDMLNIS-SLRQDGKTFIDFKKYNDKLPPLYISNP 544
 DB 1141 yylslymksekntqceltldgetlyptltkvtvnnkdnkykridlahnik-----snp 1192

RESULT 15
 AA113944 standard, Protein; 1346 AA.
 ID AA113944 standard, Protein; 1346 AA.
 XX
 XX AA113944;
 AC
 XX
 XX
 DT 15-APR-1998 (first entry)
 XX
 DE B. cereus VIPA(a)/VIP2A(a) fusion protein sequence.
 XX
 XX Vegetative insecticidal protein: Bacillus cereus strain AB78; plant;
 KM insect; Sesamia nonagrioides; maize; corn borer; toxin; fusion protein.
 OS Bacillus cereus.
 XX
 XX WO9726339-A1.
 PN
 XX 24-JUL-1997.
 PD

XX
 XX 23-DEC-1996; 96WO-EP05828.
 PF
 XX
 PR 15-JAN-1996; 96GB-0000786.
 XX
 XX
 PA (NOVS) NOVARTIS AG.
 XX
 XX
 PI Gay PB;
 XX
 XX
 DR WPI, 1997-385342/35.
 DR N-PSDB; AAT73999.
 XX
 PT Protecting plants against insects of the genus Sesamia using
 PT Bacillus toxic proteins - applied directly or expressed as
 PT heterologous protein by the plant, also transgenic plants expressing
 PT both Cry and Vip type toxins
 PS
 PS Claim 8; Page 69-75; 168pp; English.
 XX
 XX This is the amino acid sequence of a fusion protein comprising the
 CC vegetative insecticidal proteins (VIP 1a(a) and VIP2a(b)) from Bacillus
 CC cereus strain AB78. The fusion protein can be used in a new method for
 CC protecting plants, and their progeny, against insects of the genus
 CC Sesamia by direct or indirect application to the plant (or seed or
 CC growing area). The protein is especially useful to protect maize plants
 CC against the Mediterranean corn borer (S. nonagrioides).
 XX
 XX Sequence 1346 AA;

Query Match 20.5%; Score 589; DB 18; Length 1346;

Best Local Similarity 30.2%; Pred. No. 3.7e-33;
 Matches 180; Conservative 94; Mismatches 215; Indels 108; Gaps 22;

OY 1 DNLOPELKOKS-----NSRKRSTSGPTVPDRDNGIPDSLEVEGYTD 47
 DB 651 delrnpelfkkesqeflakpsklnlftqkmkreided---tdtdgsdipdlweengytl- 706
 OY 48 VKNKRRTFLSPWISNIEHKGLTKYKSSPEKWSSTASDPYSDFEKTGRIDKNVSPEARHPL 107
 DB 707 ---gnrlavkwddsl-askgytkfvsnpleshvtdydytdekaardldlsnaketfnpl 762
 OY 108 VAAVPIVHVMENIILSKNEDOSTONTDSETRTISKNTSTSRHTSEVHGNAAEVHANTST 167
 DB 763 vaafpsvnmekvllspnenls-----nsveshsstnw 796
 OY 168 SRHTSEVHGNAAEVH-----AVAIDHSLAGERTMAETMG-----LNTADTARLNNAN 215
 DB 797 syntlegasveaglgprkgsifgvsvnyqhselvaqe--wgtsqtsqfntlaasaylnan 854
 OY 216 IRVNTGTAPRYNVLPRTSLVGLKNOTLATIKAKENOLSQILAPNNVYPSKNLAPILNA 275
 DB 855 vrynnvgtalgaydvkptstfvl-ndtiatlatsnstaalinspsesyprkkgngjaits 913
 OY 276 QDDFSTPTITMANYNOFLLEKTKQLRLDQVYGNITATYFENGRRVAVDTGSNMSEVLPQ 335
 DB 914 mddfnspitltnkqvdnltnkmmletnqtdg---vykikdthgnivtggengylyq 970
 OY 336 IOETTRARIIFNGKDLNVERRIAANVPSDPLETTKPDMTKEALKIAF--GFENPGNQLQ 393
 DB 971 ikaktaslivddge-rvaekrvaakdyenpedkt-psltlkdaiklsypdeikeleglly 1028
 OY 394 YOGKDITEFDF--NFDQOTSQINIKNOIAEL-----NATNIYTVLDKIKLAKNMILIRDK 446
 DB 1029 yknkprlyessvmytidentakevtkqindtqkfkdvshlydv---klrpkmmvltk-1 1083
 OY 447 RFHYDRNNIIVAGADESVVKAHREVINSSTEG-----LLINID-----KDIR 488
 DB 1084 silydn---aesndnsigkwtntnivsgnngkkyssnmpdanlntdaqeiklnkrd 1140
 OY 489 KILSGYIVEIEDTEGLEKEVINDRYDMLNIS-SLRQDGKTFIDFKKYNDKLPPLYISNP 544
 DB 1141 yylslymksekntqceltldgetlyptltkvtvnnkdnkykridlahnik-----snp 1192

Search completed: December 2, 2001, 13:49:05
Job time: 185 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:54:52 ; Search time 132.69 seconds
(without alignments)
810.236 Million cell updates/sec

Title: US-09-747-521-4_COPY_1_735
Perfect score: 3765
Sequence: 1 MKRRKVLIPMLSTILVSS.....LYISPNKVNVAATKENT 735

Scoring table:
BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_17:*
2: sp.archaea:*
3: sp.fungi:*
4: sp.human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3649	96.9	764	2	09R0U2 bacillus an
2	3645	96.8	764	2	09KH69 bacillus an
3	3637	96.6	764	2	09F5R7 bacillus an
4	889	23.6	876	2	032739 clostridium
5	887	23.6	876	2	09KH41 clostridium
6	885	23.5	879	2	006498 clostridium
7	877.5	23.3	875	2	046221 clostridium
8	798	21.2	721	2	086171 clostridium
9	211.5	5.6	204	2	09X377 bacillus an
10	198.5	5.3	187	5	09G276 plasmidium
11	194.5	5.2	3254	5	09BR45 plasmidium
12	191.5	5.1	3130	5	09BR46 plasmidium
13	190.5	5.0	4688	2	09PC08 ureaplasma
14	187	5.0	1072	2	09CF64 lactococcus
15	183.5	4.9	2178	2	046149 clostridium
16	182	4.8	2647	5	0904X0 plasmidium
17	180.5	4.8	149	2	09RM77 clostridium
18	179.5	4.8	962	2	049546 mycoplasma
19	179.5	4.8	1051	2	049524 mycoplasma

20	179.5	4.8	1365	2	049525 mycoplasma
21	179.5	4.8	6713	2	099054 staphylococ
22	179	4.8	769	2	099V70 staphylococ
23	178.5	4.7	149	2	09RM78 clostridium
24	176.5	4.7	149	2	09RM80 clostridium
25	176	4.7	1193	2	045914 clostridium
26	174.5	4.6	149	2	09RM79 clostridium
27	174.5	4.6	4919	5	09ZHL0 haemophilus
28	173.5	4.6	2269	5	026223 plasmidium
29	173	4.6	1302	2	049547 mycoplasma
30	172.5	4.6	2747	5	09B0X9 plasmidium
31	172	4.6	1193	2	P71107 clostridium
32	172	4.6	1939	5	025662 plasmidium
33	171.5	4.6	2529	2	025579 helicobacte
34	170.5	4.5	1837	3	074424 schizosacch
35	170	4.5	1315	2	086488 staphylococ
36	169.5	4.5	1621	5	024984 giardia lam
37	168.5	4.5	2522	5	026216 plasmidium
38	168	4.5	2522	5	077365 plasmidium
39	168	4.5	3724	5	077320 plasmidium
40	167.5	4.4	1176	2	09F231 helicobacte
41	166.5	4.4	1115	3	094488 schizosacch
42	166.5	4.4	1230	2	025772 helicobacte
43	166.5	4.4	1344	2	049545 mycoplasma
44	166	4.4	604	5	026021 plasmidium
45	166	4.4	821	3	008581 saccharomyc

ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	764 AA.
09R0U2	09R0U2			
AC	09R0U2			
DT	01-MAY-2000 (TREMURel. 13, Created)			
DT	01-MAY-2000 (TREMURel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMURel. 17, Last annotation update)			
DE	PX01-110 (PROTECTIVE ANTIGEN).			
GN	PAG.			
OS	Bacillus anthracis.			
OG	Plasmid virulence plasmid PX01, and plasmid PX01.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1392;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=STERNE; PLASMID=VIRULENCE PLASMID PX01;			
RX	MEDLINE=9445483; PubMed=10515943;			
RA	Okimaka R.T., Cloud K., Hampton O., Hoffmaster A.R., Hill K.K.,			
RA	Keim P., Koehler T.M., Lanke G., Kumano S., Mahillon J., Manter D.,			
RA	Marlnez Y., Rieke D., Svensson R., Jackson P.J.;			
RT	"Sequence and organization of px01, the large Bacillus anthracis			
RT	plasmid harboring the Anthrax toxin genes.";			
RL	J. Bacteriol. 181:6509-6515(1999).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=33, 28, AND BA1035; PLASMID=PX01;			
RC	STRAIN=33, 28, AND BA1035; PLASMID=PX01;			
RX	MEDLINE=99214082; PubMed=10197996;			
RA	Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;			
RT	"Genetic diversity in the protective antigen gene of Bacillus			
RT	anthracis.";			
RL	J. Bacteriol. 181:2358-2362(1999).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=33, 28, AND BA1035; PLASMID=PX01;			
RA	Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;			
RT	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF065404; AND32414.1; -			
DR	EMBL: AF306781; AM624449.1; -			
DR	EMBL: AF306778; AM624448.1; -			
DR	EMBL: AF306779; AM624447.1; -			
DR	EMBL: AF306780; AM624448.1; -			

DR HSP: P13423; IACC.
DR InterPro: IPR003896; Binary_toxB.
DR PRINTS: PR01391; BINARYTOXB.
KW Plasmid.
SQ SEQUENCE 764 AA; 85810 MW; 3AE1EFBA48FA03F CRC64;

Query Match 96.9%; Score 3649; DB 2; Length 764;
Best Local Similarity 97.4%; Pred. No. 4.7e-204;
Matches 716; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

OY 1 MKRRVLPPLMALSTILVSTGNLEVIOAEVQKOEKRLNESSSSGGLGYYFSDLNFOA 60
DB 1 MKRRVLPPLMALSTILVSTGNLEVIOAEVQKOEKRLNESSSSGGLGYYFSDLNFOA 60
OY 61 PMVVTSTTGDISIPSSSELENIPSENOYFQSAIWSGFIKKKSDEYFATSADNHTVMV 120
DB 61 PMVVTSTTGDISIPSSSELENIPSENOYFQSAIWSGFIKKKSDEYFATSADNHTVMV 120
OY 121 DDOEVINKASNSNKIRLEKGRLYQIKIOYORENPTEKGLDFLWYTDSONKKEVISSDNL 180
DB 121 DDOEVINKASNSNKIRLEKGRLYQIKIOYORENPTEKGLDFLWYTDSONKKEVISSDNL 180
OY 121 DDOEVINKASNSNKIRLEKGRLYQIKIOYORENPTEKGLDFLWYTDSONKKEVISSDNL 180
DB 121 DDOEVINKASNSNKIRLEKGRLYQIKIOYORENPTEKGLDFLWYTDSONKKEVISSDNL 180
OY 181 QLPETLKOKSSNRKRSTASGPTVPDRNDGIPDSLEVEGYVDVKNKRTFLSPWISNH 240
DB 181 QLPETLKOKSSNRKRSTASGPTVPDRNDGIPDSLEVEGYVDVKNKRTFLSPWISNH 240
OY 241 EKKGLTKYKSSPEKWSSTADPYSDFEKYTGRIIDKNVSPRARHPVLAAYPIVHVDMENIL 300
DB 241 EKKGLTKYKSSPEKWSSTADPYSDFEKYTGRIIDKNVSPRARHPVLAAYPIVHVDMENIL 300
OY 301 SKNEOSTONTDSETRITISKNTSTSRTHSEVHGNAEVAHANTSTSRTHSEVHGNAEVAH 360
DB 301 SKNEOSTONTDSETRITISKNTSTSRTHSEVHGNAEVAHANTSTSRTHSEVHGNAEVAH 360
OY 361 VAIDHSLSLAGRTAEFTGLTADTARLANANIRVYNTGTAPVLTPTSLVLKNOFL 420
DB 361 VAIDHSLSLAGRTAEFTGLTADTARLANANIRVYNTGTAPVLTPTSLVLKNOFL 420
OY 421 ATIKAKENOLSOILAPNNYPSKNIAPALNADDFSSPTITMANYNOFLELEKTRQLRD 480
DB 421 ATIKAKENOLSOILAPNNYPSKNIAPALNADDFSSPTITMANYNOFLELEKTRQLRD 480
OY 481 TDOYGNATYTFENGRAVVDGSMSEVLPQIOETTAIIENGKDLNVERRIAIVNPS 540
DB 481 TDOYGNATYTFENGRAVVDGSMSEVLPQIOETTAIIENGKDLNVERRIAIVNPS 540
OY 541 DPLETTKPDMTLEKALIFGFENEPNGNLQYOGKDIPEDFNFDOOTSNIKOLAELNA 600
DB 541 DPLETTKPDMTLEKALIFGFENEPNGNLQYOGKDIPEDFNFDOOTSNIKOLAELNA 600
OY 601 TNIYTVLDKIKLNAKKNILIRDKRFHYDRNNITAVGADSVKAEHREVINSSTEGILLNI 660
DB 601 TNIYTVLDKIKLNAKKNILIRDKRFHYDRNNITAVGADSVKAEHREVINSSTEGILLNI 660
OY 661 DKDIIKILSGYIVLEDETEGLKEVINDRYDMLNISLRODGKTFIDFKKYNKPLYSIN 720
DB 661 DKDIIKILSGYIVLEDETEGLKEVINDRYDMLNISLRODGKTFIDFKKYNKPLYSIN 720
OY 721 PMYKVNVAVTRENT 735
DB 721 PMYKVNVAVTRENT 735

RESULT 2

O9KH69 PRELIMINARY; PRT; 764 AA.
AC O9KH69;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PROTECTIVE ANTIGEN.
GN PACA OR PAG.

OS Bacillus anthracis.
OG Plasmid pX01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxId=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-V770-NP1-R, ATCC14185, PLASMID-PX01;
RA MEDLINE=20359347; PubMed=10899854;
RX Cohen S., Mendelson I., Alboum Z., Kohler D., Elhanany E., Bino T.,
RA Leitner M., Inbar I., Rosenberg H., Gozes Y., Barak R., Fisher M.,
RA Kromann C., Velan B., Shaffer A.,
RT "Attenuated nontoxicogenic and nonencapsulated recombinant Bacillus
RT anthracis spore vaccines protect against anthrax".
RL Infect. Immun. 68:4549-4558(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC PLASMID-PX01;
RA MEDLINE=99214082; PubMed=10197996;
RX Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RT "Genetic diversity in the protective antigen gene of Bacillus
RT anthracis".
RL J. Bacteriol. 181:2358-2362(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC PLASMID-PX01;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF268967; AAF6457.1; -
DR EMBL: AF306782; AAG24450.1; -
DR InterPro: IPR003896; Binary_toxB.
DR PRINTS: PR01391; BINARYTOXB.
KW Plasmid.
SQ SEQUENCE 764 AA; 85838 MW; BDDAFCDCE83DAF4 CRC64;

Query Match 96.8%; Score 3645; DB 2; Length 764;
Best Local Similarity 97.3%; Pred. No. 8.1e-204;
Matches 715; Conservative 3; Mismatches 17; Indels 0; Gaps 0;

OY 1 MKRRVLPPLMALSTILVSTGNLEVIOAEVQKOEKRLNESSSSGGLGYYFSDLNFOA 60
DB 1 MKRRVLPPLMALSTILVSTGNLEVIOAEVQKOEKRLNESSSSGGLGYYFSDLNFOA 60
OY 61 PMVVTSTTGDISIPSSSELENIPSENOYFQSAIWSGFIKKKSDEYFATSADNHTVMV 120
DB 61 PMVVTSTTGDISIPSSSELENIPSENOYFQSAIWSGFIKKKSDEYFATSADNHTVMV 120
OY 121 DDOEVINKASNSNKIRLEKGRLYQIKIOYORENPTEKGLDFLWYTDSONKKEVISSDNL 180
DB 121 DDOEVINKASNSNKIRLEKGRLYQIKIOYORENPTEKGLDFLWYTDSONKKEVISSDNL 180
OY 121 DDOEVINKASNSNKIRLEKGRLYQIKIOYORENPTEKGLDFLWYTDSONKKEVISSDNL 180
DB 121 DDOEVINKASNSNKIRLEKGRLYQIKIOYORENPTEKGLDFLWYTDSONKKEVISSDNL 180
OY 181 QLPETLKOKSSNRKRSTASGPTVPDRNDGIPDSLEVEGYVDVKNKRTFLSPWISNH 240
DB 181 QLPETLKOKSSNRKRSTASGPTVPDRNDGIPDSLEVEGYVDVKNKRTFLSPWISNH 240
OY 241 EKKGLTKYKSSPEKWSSTADPYSDFEKYTGRIIDKNVSPRARHPVLAAYPIVHVDMENIL 300
DB 241 EKKGLTKYKSSPEKWSSTADPYSDFEKYTGRIIDKNVSPRARHPVLAAYPIVHVDMENIL 300
OY 301 SKNEOSTONTDSETRITISKNTSTSRTHSEVHGNAEVAHANTSTSRTHSEVHGNAEVAH 360
DB 301 SKNEOSTONTDSETRITISKNTSTSRTHSEVHGNAEVAHANTSTSRTHSEVHGNAEVAH 360
OY 301 SKNEOSTONTDSETRITISKNTSTSRTHSEVHGNAEVAHANTSTSRTHSEVHGNAEVAH 360
DB 301 SKNEOSTONTDSETRITISKNTSTSRTHSEVHGNAEVAHANTSTSRTHSEVHGNAEVAH 360
OY 361 VAIDHSLSLAGRTAEFTGLTADTARLANANIRVYNTGTAPVLTPTSLVLKNOFL 420
DB 361 VAIDHSLSLAGRTAEFTGLTADTARLANANIRVYNTGTAPVLTPTSLVLKNOFL 420
OY 421 ATIKAKENOLSOILAPNNYPSKNIAPALNADDFSSPTITMANYNOFLELEKTRQLRD 480
DB 421 ATIKAKENOLSOILAPNNYPSKNIAPALNADDFSSPTITMANYNOFLELEKTRQLRD 480
OY 481 TDOYGNATYTFENGRAVVDGSMSEVLPQIOETTAIIENGKDLNVERRIAIVNPS 540

```
Db 481 TDQVYGNIAATYFNENGRVAVDTGSMNSEVLPOIQTETARIIFNGKDLNVERRIAANPS 540
Qy 541 DPLETTKPDMTLKEALKIAGFNEPNCNLOYGKDIETEPFNFDOOTSQNIKNOLAELNA 600
Db 541 DPLETTKPDMTLKEALKIAGFNEPNCNLOYGKDIETEPFNFDOOTSQNIKNOLAELNV 600
Qy 601 TNYTTLVDKIKLNKAMNILLRDKRFHYDRNNIAVGADESYYKEAHRVINSSTEGLLNI 660
Db 601 TNYTTLVDKIKLNKAMNILLRDKRFHYDRNNIAVGADESYYKEAHRVINSSTEGLLNI 660
Qy 661 DKDIRKILSGYIEIEDTEGLKEVINDRYDMLNLSLRDQKTFIDFKKYNDKLPYISN 720
Db 661 DKDIRKILSGYIEIEDTEGLKEVINDRYDMLNLSLRDQKTFIDFKKYNDKLPYISN 720
Qy 721 PNKYVNVYAVTKENT 735
Db 721 PNKYVNVYAVTKENT 735

RESULT 3
Q9FSR7 PRELIMINARY; PRT; 764 AA.
AC Q9FSR7;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE PROTECTIVE ANTIGEN.
GN PAG.
OS Bacillus anthracis.
OC Plasmid pX01.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Bacillus.
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BA1024;
RX MEDLINE=99214082; PubMed=10197996;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RT "Genetic diversity in the protective antigen gene of Bacillus anthracis."
RT J. Bacteriol. 181:2358-2362(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BA1024.
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF306783; AAG24451.1; -.
DR InterPro; IPR003896; Binary_toxb.
DR PRINTS; PR01391; BINARTOXINB.
KW plasmid.
SQ SEQUENCE 764 AA; 85828 MW; A1845CE1FEDCD93A CRC64;
```

Query Match 96.6%; Score 3637; DB 2; Length 764;
Best Local Similarity 97.1%; Pred. No. 2.4e-203;
Matches 714; Conservative 3; Mismatches 18; Indels 0; Gaps 0;

```
Qy 241 EKKGLTKYKSSPEKWSASDPYSDFEKVGTGRIDKNSPEARHPLVAAVPIVHVMENITL 300
Db 241 EKKGLTKYKSSPEKWSASDPYSDFEKVGTGRIDKNSPEARHPLVAAVPIVHVMENITL 300
Qy 301 SKNEDQSTQMTDSEPTISKNTSFRTHTESEVHNAEVAHNTSTSRTHTESEVHNAEVAH 360
Db 301 SKNEDQSTQMTDSEPTISKNTSFRTHTESEVHNAEVAHNTSTSRTHTESEVHNAEVAH 360
Qy 361 VAIDHSLSLAGEERTWAETMGLNTADTARLNANIRYVNTGTAPIYVNLPTTSVLGKQOTL 420
Db 361 VAIDHSLSLAGEERTWAETMGLNTADTARLNANIRYVNTGTAPIYVNLPTTSVLGKQOTL 420
Qy 421 ATTKAKENQSLIAPANNYPKSLAPALNAQODESSPTPTMAYNPFLEKTKOIRLD 480
Db 421 ATTKAKENQSLIAPANNYPKSLAPALNAQODESSPTPTMAYNPFLEKTKOIRLD 480
Qy 481 TDQVYGNIAATYFNENGRVAVDTGSMNSEVLPOIQTETARIIFNGKDLNVERRIAANPS 540
Db 481 TDQVYGNIAATYFNENGRVAVDTGSMNSEVLPOIQTETARIIFNGKDLNVERRIAANPS 540
Qy 541 DPLETTKPDMTLKEALKIAGFNEPNCNLOYGKDIETEPFNFDOOTSQNIKNOLAELNA 600
Db 541 DPLETTKPDMTLKEALKIAGFNEPNCNLOYGKDIETEPFNFDOOTSQNIKNOLAELNV 600
Qy 601 TNYTTLVDKIKLNKAMNILLRDKRFHYDRNNIAVGADESYYKEAHRVINSSTEGLLNI 660
Db 601 TNYTTLVDKIKLNKAMNILLRDKRFHYDRNNIAVGADESYYKEAHRVINSSTEGLLNI 660
Qy 661 DKDIRKILSGYIEIEDTEGLKEVINDRYDMLNLSLRDQKTFIDFKKYNDKLPYISN 720
Db 661 DKDIRKILSGYIEIEDTEGLKEVINDRYDMLNLSLRDQKTFIDFKKYNDKLPYISN 720
Qy 721 PNKYVNVYAVTKENT 735
Db 721 PNKYVNVYAVTKENT 735

RESULT 4
ID 032739 PRELIMINARY; PRT; 876 AA.
AC 032739;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE ADP-RIBOSYLTRANSFERASE.
GN CDTB.
OS Clostridium difficile.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD196;
RX MEDLINE=97230316; PubMed=9119480;
RA Perelle S., Gilbert M., Bourlioux P., Corthier G., Popoff M.R.;
RT "Production of a complete binary toxin (actin-specific ADP-ribosyltransferase) by Clostridium difficile CD196."
RT Infect. Immun. 65:1402-1407(1997).
DR EMBL; U76081; AAC67305.1; -.
DR HSSP; P13423; AAC.
DR InterPro; IPR003896; Binary_toxb.
DR PRINTS; PR01391; BINARTOXINB.
KW transferase.
SQ SEQUENCE 876 AA; 98796 MW; 25E06E2D45CE2B3B CRC64;
```

Query Match 23.6%; Score 889; DB 2; Length 876;
Best Local Similarity 32.5%; Pred. No. 1.3e-43;
Matches 261; Conservative 139; Mismatches 274; Indels 128; Gaps 34;

DT 01-JUL-1997 (TREMblrel. 04, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE SB COMPONENT.
GN SBL.
OS Clostridium spiroforme.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes.
OX NCBI_Taxid=29348;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CS246;
RA Gilbert M., Perelle S., Daube G., Popoff M.R.;
RL Syst. Appl. Microbiol. 20:337-347(1997).
DR EMBL: X97699; CAA66612.1; -.
DR HSSP: P13423; IACC.
DR InterPro: IPR003896; Binary_toxB.
DR PRINTS: PRO1391; BINARYTOXINB.
SQ SEQUENCE 879 AA; 98738 MW; 40685ACB8E05BA01 CRC64;

Query Match 23.3%; Score 885; DB 2; Length 879;
Best Local Similarity 29.0%; Pred. No. 2,2e-43;
Matches 267; Conservative 148; Mismatches 269; Indels 236; Gaps 36;

OY 1 MKRKVLIPLMALSTLVSTGNLEVIQAEV-----KQENRLNSESSESSOGLLGYFSD 55
DB 1 MNKRLTGLTLC--TVVGQMMTPYAKTIQNTQNDQEVETNEKTVSSNGMLGYEAD 58
OY 56 LNFQAPMVYTSSTGDLSPSSELENIPSENO-YFQSAIWSGFYKKSDEYFATSADN 114
DB 59 EHFKEDELKAPYKNGELKEKKNVEKLTLEEKTNIKSIRHTGRIPSKDEYTLSTDKN 118
OY 115 HTVMVDDQEVINKASNSKIRLEKRLQIKYQRENTENGDLF-----KLYMTD 167
DB 119 -VLMQINAE--GEIANTPLKNNIKQGEYSIRLEIQ-----DKDIGVDDLSPKLYW-E 168
OY 168 SONKEVSSDMLPELQKSSN-----SRKRSTSA-GPYPPDRDNDG 211
DB 169 LMGDKTLIEKKNFLRDYKIDENDPFIKDNFDLKLRSARLASGWCDELDLDNDN 228
OY 212 IPDSLEVEGYVDVKNKRTFLSPWISNIHEKGLTYKSSPEKWSPASDYSDFEYVGR 271
DB 229 IPDAYEKNQYTI---KDSIAVYK-W-EDSFAQGGYKKRYLSYSLESTNAGDPYDYQKASG 283
OY 272 IOKNVSPERHPLVAAVPIYHDMMENILSKNDOSTQNTDSTRTISKTSRTP--HT 329
DB 284 FKAIAEARDPLVAAVPAVVGOMKELIISTNEHAST---DQKTYVSRTNTSKTDANT 339
OY 330 SEVHGNAEVH---ANTSTSRTHSEVHGNAEVAHVAIDHSLSLAGEPTMAETMGLNTAD 385
DB 340 AGVAIINAVQNGFTGSIITNYSHTTE-----NSTAVONS---NGE-SMNTSLSTIKGE 388
OY 386 TARLANINRYVTGTAPIYVNLPTSTVLGKNOTLATIRAKENQLSQILAPNNYSPSKNL 445
DB 389 SAYINANVRYVTGTAPIYKVPPTNLYL-DGDTLTITKAQNDQINNLSPNTYIRKKGL 447
OY 446 APLALNAODESSPTITMANYNQFLELEKTKQLDLTDYQYGNATATNFENGRRVDT-GS 504
DB 448 SPLALNTMOQFSSRLPIRYNDOLKIDAGKOIKLETTQVSGN---YGIKNSQOQIITEGN 504
OY 505 MNSEVLPOIOWETARIIFN-GKDLNVERRIAIVNSDPLETTKPRMTLKEAKIKFGEN 563
DB 505 SMSDYISQDLSASITLDTGSD--VFERRVYAKDSSNEDKTP-PVLTIGEALEKKAFGAT 561
OY 564 EFGNGLQYQKQDTE--FDENFDQOSTONIKQNLQALNATNTIYVLDKIKLNKMNILR 621
DB 562 KNGEILYFGMPIDESCVELIFDGNANTLIKERLNLNDKIKYINV---QLEGMKMLIK 617
OY 622 D----- 622
DB 618 TSTYFNNDGYNNFPSSWSNVDSDNODGLQANANKLGETKIVIPSKLNPYKRYVFSGY 677
OY 623 -----KRFHY-----DRNN----- 621

DB 678 LKNSSTNSPITVINIKAKEQTYNLVSENDYKKSSEYEFETIGRPASNIETILTSSGITFLD 737
OY 632 -----IANGADESVYKKAHREV-----INSTDELLN-----I 660
DB 738 NLSITELNSTPELKEPDIIPSDQELI-DAHKRYVADLSFNOSTANYIYDGLYFEPQOT 796
OY 661 DKDIRKLSGYIEIE-DTEGLKEVINDRYDMLNISSLRDQGT-FIDFKYNDKLPYLI 718
DB 797 NKEVLDYQKRYEATIEYSGFKDIGTKKELRNRYTDSNQPTNYNFRSY-----FT 850
OY 719 SNPNY-----KVNYAVATKEN 734
DB 851 SGENVMPYKKRLRIYATIPEN 870.

RESULT 7
ID 046221 PRELIMINARY; PRT; 875 AA.
AC 046221;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE IOTA TOXIN COMPONENT IB PRECURSOR.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OX NCBI_Taxid=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCIB 10748;
RX MEDLINE=94041637; PubMed=8225592;
RA Perelle S., Gilbert M., Bogue P., Popoff M.R.;
RT Characterization of Clostridium perfringens iota-toxin genes and
expression in Escherichia coli.*
RL Infect. Immun. 61:5147-5156(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-NCIB 10748;
RA Popoff M.R.;
RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-NCIB 10748;
RA Popoff M.R.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: X73562; CAA51960.1; -.
DR HSSP: P13423; IACC.
DR InterPro: IPR003896; Binary_toxB.
DR PRINTS: PRO1391; BINARYTOXINB.
KW Signal.
FT SIGNAL 34 38
FT CHAIN 212 875 IOTA TOXIN COMPONENT IB.
SQ SEQUENCE 875 AA; 98468 MW; C9AE092CD3818921 CRC64;

Query Match 23.3%; Score 877.5; DB 2; Length 875;
Best Local Similarity 31.9%; Pred. No. 5,8e-43;
Matches 254; Conservative 132; Mismatches 270; Indels 139; Gaps 33;

OY 15 TLIVSSGNLEV-----IOAEVQENRLNSESSESSOGLLGYFSDNFPQAPMVYTSST 69
DB 15 TAMIQSLSTNVYAQTTOQDNTQKEETINENTLSSNGMLGYFADHEFDLELMADIKN 74
OY 70 GDLSPSSELENIPSE-NQYFQSAIWSGFYKKSDEYFATSADNHTVMVDDQEVINK 128
DB 75 GDLKFEKKYDKLLETDENSSIKSIRWTGRILIPSDGYILSTDR-NVLMQINAKGDIK 133
OY 129 ASNSNKIRLEKGRLYQIKIYOREN-----PTKGLDFKLYWTDSONKKEVSSDN 179
DB 134 ---TLKVNMRKGQAYNIRIETIODKNGSIDNLSVP-----KLYW-ELMGKNTVPIPEN 182
OY 180 LQLELQKSSNSRKRKSTSGPTVP-----DNDNGIJDLSLEVG 220

```
Db 183 LPRFRYSKIDEDD-----PPIPNNNPFDVREFSSAAMEDDEDIDTDNDNIPDAVERNG 233
Qy 221 YTVVKKNRKRLSPWISNIHEKKGLTKYSSPEKSTASDPYDEKATGRIDKVVSPA 280
Db 224 YTI-----KDSIAVKNNDSEFAE-QGYKKYVSSLESNTAGDPYTDQKASGSIDKAKIKLEA 288
Qy 281 RHPLYAAPIYHVDMENTILSKNEDOSTONTDSETRTISKNTSRTSHTEVHGNAEYHA 340
Db 289 RDPLYAAPIYVGVGMENILISTNEHASS-----DQKTVSRATTNKNTQANTV--GVSTISA 342
Qy 341 -----NTGSTRHTESEVHGNAEYHVAIDHSLSLAGEPTMATMNLADTARLAN 392
Db 343 GYNGFTGNITTSYSHTD-----NSTAVODS---NGE-SMWTGLSINGESAYIVAN 391
Qy 393 IRYVNTGAPRIYNYLPTTSYLKGNQTLATIKAKENOLSOILAPNNYPSKMLAPIALNA 452
Db 392 VRYNTGAPRYKYTPNTNLV-DEETLATIKADONQIGNNISPEMYTPKKGSLPLAINT 450
Qy 453 ODDFSSPTIYNNVNOFLEKTKOIRLDTDOYGNIAIYFNENGRVVDGSSNMSEVLPO 512
Db 451 MDOFNARLPIYNYDOLKLDGSKQIKIETTVQSGNYGTKN--SQGI-IIEGNSMGNYISQ 508
Qy 513 IQETTARIIFNGKOLNVERIAANPSDPLETKPDMTLKALKIAGFNEPNKNLOY- 571
Db 509 IDSVASITILD-TGSQTFERRVAAKEQGNPEDKT-PEITIGELAKKAFSAFK-NGELLYF 565
Qy 572 QGKDITE--PDFNPDQOSTQNIKNOLAELNATNYTVLDKIKLNKMLIRDRKF--HY 627
Db 566 NGIPIDECVELIFDDNISEIIEKQLKYLDDOKKIYV-----KLENGMILLIKVSPYTFNF 621
Qy 628 DR-NNIAVGADESVYKKAHREYINSTEGL-----LNIIDKIRKILSGY----- 671
Db 622 DEYNNFP--ASMSNIDFTNODGLQSVANKLSGEFKIIPMSKIKPKYKRYVFSKDPST 679
Qy 672 -----IVEIEDREGLEKEVYNDRYDMNLIS-----SLROGKTFIDPKKND 712
Db 680 SNSTIVNIKSEQKTDYLVPEKDYTKFSYEFETTGKSSDIETLTSSGVIFLDMLSITE 739
Qy 713 --KLPLYSINPNKYV 725
Db 740 LNSTPELKEPEIKV 754

RESULT 8
086171 PRELIMINARY; PRT; 721 AA.
AC 086171:
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE C2 TOXIN (COMPONENT-11).
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TYPE C (C)-203028;
RX MEDLINE=98323874; PubMed=9659689;
RA Kimura K., Kubota T., Ohishi I., Isogai H., Isogai E., Fujii N.;
RT "The gene for component-II of botulinum C2 toxin."
RL Vet. Microbiol. 62:27-34(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-TYPE C (C)-203028;
RX MEDLINE=96184657; PubMed=8645309;
RA Fujii N., Kubota T., Shitake S., Kimura K., Ohishi I., Morishiki K.,
RA Isogai E., Isogai H.;
RT "Characterization of component-I gene of botulinum C2 toxin and PCR
RT detection of its gene in clostridial species."
RL Biochem. Biophys. Res. Commun. 220:353-359(1996).
DR EMBL; D88982; BAA32537.1; -
```

```
DR HSSP; P13423; IACC.
DR InterPro; IPR003896; Binary_toxB.
DR PRINTS; PRO1391; BINARYTOXINB.
SQ SEQUENCE 721 AA; 80515 MW; 44C8153AC749D5F2 CRC64;

Query Match 21.28; Score 798; DB 2; Length 721;
Best Local Similarity 34.58; Pred. No. 1.86-38;
Matches 221; Conservative 117; Mismatches 221; Indels 82; Gaps 25;

Qy 32 KOENRLNESES--SSOGLLGYPSSDLNFOAPMVYTSSTDDGISPSELEINIPSENYF 89
Db 5 KPENSVKNNKNYFTINGLMGYFEN-DFENLNTISPLDNLIFSKEDINSTIG-NKTI 62
Qy 90 QSAISGFIKVKSDSEYFATSADN-HVTMWVDDQEVINKASN--SNKIRLEKGRLYQIKI 147
Db 63 KSARIGIKLKSITGEIYLLTNSPCREL---NEIENLSLNSNTYVNLQGNVYDIRI 119
Qy 148 -QYORENPTKGLD-FKLYWTDSONKKEVISSDNLQPELKOKSSNRKSTSGPTV- 204
Db 120 EQLMSENOLLNKYEGIKLYWETSDIKEIIPSEVL---LKPNSYNTNEKSKFIPNNTLF 175
Qy 205 -----PDRONDGIPTDSELEVGYVDYKNNKRTFLSPMISNIHEKKGLTKYSSPEK 254
Db 176 SNAKIKANANDTDGDIPEDEWEINGYVMNOKAVAMDCKEPAAN-----GKKYVSNPFK 230
Qy 255 WSTASDPYSEFEKYGRIDKNVSPPEARHPLVAAPIYHVDMENTILSKNEDOSTONTDSE 314
Db 221 PCTANDPYTDFEKVSGQIDPSVSWARDPMSIATYIVGVOMERLVVKSFE---YITDSD 286
Qy 315 TRTISKNTSRTHT---SEVHGNAEV-----HANTSTSKTHSEVHGNAEV 358
Db 287 TKSMSKSTSHSTNTNITGAEVSGSLQAGLIFPVFSMASANSYHTQMNSTVD----- 341
Qy 359 HAVAIDHSLAGEPTMAETMGLNADTARLANRNYNTGAPRIYNYLPTTSYLKKNQ 418
Db 342 -----DTTGE-SFSQGLSINGESAYINPNIRYNTGAPRYVNTPTTIVTDK-Q 390
Qy 419 TLATIKAKENOLSOILAPNNYPSKMLAPIALNADDESPITMNNVNOFLEKTKOLR 478
Db 391 SVATIKGQESLIGDVLNCGTYPIIGEPFMALNTMDQSSFLPIYVNLQKSIDCGTYM 450
Qy 479 LPTDOYVGNIAIYFNENGRVVDGSSNMSEVLPOIETTARI--ENKGLNLYERRIAA 536
Db 451 LSTSQFTGNFAKYN-SNGLNLYTD-GNNMGPLYGITIKSTYASLTLSFSQOTVOVA---VVA 505
Qy 537 VNPSPLETTKPDMTLKALKIAGFNEPNKNLOQOGDII---EFDNPDQOSTQNIKN 593
Db 506 PNFSPDEKTK-PLTLLEQALYKAFALKKNGKFFHGLEISKNEKIOVFLDSNTNDFEN 564
Qy 594 QLAELNATNYTVLDKIKLNKMLNLIIRDRFHYDRNIAV 634
Db 565 QLNKTAADKDINHCI--IKRN--MNLVAVITFKENISSINI 601

RESULT 9
09X377 PRELIMINARY; PRT; 204 AA.
AC 09X377:
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, last annotation update)
DE PX01-111.
OS Bacillus anthracis.
OC Bacterium; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-STERNE;
RA Ohtnaka R.T., Cloud G., Hamton O., Hoffmaster A., Hall K.K., Keim P.,
RA Koehler T., Lamke G., Kumano S., Mahillon J., Mantler D., Martinez Y.,
```

RA Rieke D.O., Svensson R., Jackson P.J.;
RT "the sequence and organization of pXO1, the large Bacillus anthracis
RT plasmid harboring the Anthrax toxin genes."; J. Bacteriol. 0:0-0(1999).
RL EMBL: AF065404; AAD32415.1; -.
DR HSSP: P13423; IACC.
KW plasmid.
SO SEQUENCE 204 AA; 23029 MW; E1657B23AE4273FD CRC64;

Query Match	5.68;	Score 211.5;	DB 2;	Length 204;
Best Local Similarity	34.88;	Pred. No. 3.7e-05;		
Matches 46;	Conservative 31;	Mismatches 34;	Indels 21;	Gaps 4

OY 676 EDTE-----GLKEVINDRYDAMINISLRDDGKTFIDDKFKNVONKLPLYLSNPN 722

Dd 60 KKSBNHLTNSPVTTTLAGKDSGVVELRVLS-----DGTFIDFKFNKDENMSLV-DPG 112

OY 723 YKANVYAATKEN 734

Dd 113 DDVYVAATKED 124

	RESULT	10	
Q9GZ76			
ID	Q9GZ76	PRELIMINARY;	PRT; 1387 AA.
AC	Q9GZ76;		
DT	01-MAR-2001 (TREMBLE)	16,	(Created)
DT	01-MAR-2001 (TREMBLE)	16,	Last sequence update)
DT	01-MAR-2001 (TREMBLE)	16,	Last annotation update)
DE	RETICULOCYTE-BINDING PROTEIN 2 HOMOLOG A (FRAGMENT).		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5833;		
RX	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20402589; PubMed=10920203;		
RA	Raeyner J.C., Gallinski M.R., Ingravallo P., Barnwell J.W.;		
RT	"Two Plasmodium falciparum genes express merozoite proteins that are		
RT	related to Plasmodium vivax and Plasmodium yoelii adhesive proteins		
RT	involved in host cell selection and invasion."		
DR	Proc. Natl. Acad. Sci. U.S.A. '97:9648-9653(2000).		
RL	EMBL: AF196347; AAF98066.1; -.		
FT	NON_TER	1	1
FT	NON_TER	1387	1387
SQ	SEQUENCE	1387 AA;	164871 MW; 5660154744FQC285 CRC64;

Query Match	5.38;	Score 198.5;	DB 5;	Length 1387;
Best Local Similarity	19.68;	Pred. No. 0.0031;		
Matches 169;	Conservative 136;	Mismatches 289;	Indels 269;	Gaps 41

Oy	5	KVLPLMALSTIIIVSSNGNLEVIQAEVKOEKREINSESSOGGLGYGFSLNPQAAMV	64
Dd	284	KHVIKLANSFIIIMSDTNETIITPENPLEDDLLN-----LQILFERKEHEI	329
Oy	65	TTSPTTGDLSTPSSLEENI--PSENQYFOSAIWGCFIKVAKSDEYTFATSADNHVTMVD	122
Dd	330	TSTLENSDLS---ELDHIGSNSDESIDNLKVYNNIEFL-----HTYST---	QLKYLDN
Oy	123	-----QEVIKASN-SKKITPLEGRLYQT-KIQ	148
Dd	378	IQLKGDCNDLVKDCKELRELSTALYDLKIOTYSVINREDNISNNIDIVSKNLNEIDAIO	437
Oy	149	YORE-----NPETEKGLPFKLYWTDSQMKKE	177
Dd	438	YNFEKYKEIPDNVEEYKTLLDTKNAYIVYKKAELKANVDINKTKRDEDLIYFNDELKESL	497
Oy	174	VISSNDLOLPCLKOKSSNRKRKSTSAGPTVPDRDNDG---TP---DSLWEGYTVD----	224

```

Db 498 TLSSNEMEIKTYONSVNS-----FSDJKNKINDIDEMKTLIPMDLLENEHNIDSLY 553
Oy 225 ---VKNRTELSPWISNIEHKGLT-----KYKSPKRWSTAS---DPYSDEKYVTGRID 273
Db 554 NFIIIRNOIKIGINDIKNIREOENDTNICEFYIONNNNPFKSDISIFNKYDHIKVDNIS 613
Oy 274 KNSPEARHPPLVAAYRPIVAYHD--MENI-----IJSKNED-----OSNO-----NTQSE 314
Db 614 NNIDVVKHSHLSSEHYINNTIENIMSYIVAINEDTEMSNLEETQDLLELYENPKKE 673
Oy 315 TRTISKMTSTSR-THTSEVHGNAEVAHANTSTPHTSEVHG-----MAEV 358
Db 674 KNIIINNKKIYHFENKLEKEIENSLETYSNISTENFKINETOINIDILKNEENNITKINDKV 733
Oy 359 -HAVAIDHSLSLAGEPERWAETMG---LNPADTARL-----NANIRVYNTGAPIVLPTT 410
Db 734 KELYHVSTLTLESTIQFNNULYGDLSNIDOVKYKEDIYNVELAKKYKLYIENITHLGRJ 793
Oy 411 SLVLGKNQTLATTIKAKENQLS---QILAPNNYPSKNLAPIALNADDESSPTITMANYQ 467
Db 794 NTFI---KELDKYODENNGIDKYIETIKENNSYTIK-LKEKANLNKENIS-----K 840
Oy 468 FLELEKTKOLRLTDQYUGNATLYNENGVRADTOSMNSSEVLPOIQETARTIIFNGDL 527
Db 841 LLQOIKRNETEL-----YNNIN--IKD-----IANTGASV 869
Oy 528 NLVERRIAAYNPSDPLETTKPDMTLKEALKIARGFNEPNGNLOYOGKDIITEPDFNFDOOT 587
Db 870 NNIKQFSS-----NLPLEKEL---FQMEMLNLI----- 896
Oy 588 SQNKTKQALAEINATITYVL-----DKKLAKANNILIRPKRF-HYDRNNIAVGADES 639
Db 897 -NNIMNENKTRISNDVYNTTTLQDIENNKRENNNNNIETIKLIDHIKIHEKIQOAEIL 955
Oy 640 VVKEAHEVINSSTEGGLNIDKDIRKISGLYVE---IEDTEGLKE---YINDRYD- 690
Db 956 IIDDAKRVY-----KEITDININKAFNEITENYNNENNGYIKSAKNIVDEATYLNNELDKF 1010
Oy 691 MLNITSSLRQDGKTFIDPKKYNDK 713
Db 1011 LKLNELLSHNNN--DIKDLGDE 1031

```

RESULT	11	
Q9BK45		
ID	Q9BK45	PRELIMINARY; PRT; 3254 AA.
AC	Q9BK45;	
DT	01-JUN-2001 (TReMBLrel, 17, Created)	
DT	01-JUN-2001 (TReMBLrel, 17, last sequence update)	
DT	01-JUN-2001 (TReMBLrel, 17, last annotation update)	
DE	RETICULOCYTE BINDING PROTEIN 2 HOMOLOG B.	
OS	Plasmodium falciparum.	
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
OX	NCBI_TaxID=5833;	
RI	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=21101060; Pubmed=11160005;	
RA	Trigilia T., Thompson J., Caruana S.R., Delorenzi M., Speed T.,	
RA	Cowan A.F.;	
RT	"Identification of Proteins from Plasmodium falciparum That Are	
RT	Homologous to Reticulocyte Binding Proteins in Plasmodium vivax.";	
RL	Infect. Immun. 69:1084-1092(2001).	
DR	EMBL: AF212917; AAK19245.1; -;	
SO	SEQUENCE 3254 AA; 382876 MW; 669CAFA5A6167BA CRC64;	

```

Query Match          5.28; Score 194.5; DB 5; Length 3254;
Best Local Similarity 19.68%; Pred. NO. 0.017;
Matches 169; Conservative 135; Mismatches 290; Indels 269; Gaps 41.

QY 5 KLLIPALSTLVSTGCLVEYIQAEVKQENRLNSESSESGGLCYTSDLDNFQAPMVY 64
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

```

QY 5 KVLPLMALSTLIVSSTGNLEVIQAELVKQENRRLNSESSESSQGLLGYYFSDNLFGAPMNV 64

Db 1920 KHVILANSGIIVMSDTTEITPENPLEDNDLN-----LQLYFERKHEI 1965
Qy 65 TSSITGDLSPSELENT--PSENOYFQSAISGFIKVKKSDEYFPATSADNHVTWVVD 122
Db 1966 TSTLENDSDL---ELDHGNSDESIDNLKYVNDIEL-----HMYST---QILKYLDN 2013
Qy 123 -----OEYINKASN-SNKIRLEKGRLOYI-KIO 148
Db 2014 IOKLGDGCDLVKDKELRELSTALYDLKIOITSVINRENDISNNIDISNKLNEIDAIO 2073
Qy 149 YORE-----NPEKGLDFKLYWTDSONKKE 173
Db 2074 YNFEKYKEIFDQVVEEYKTLDDTKNAYIVKKAELIKVNDINKREDIDYFNDLEDEKSL 2133
Qy 174 VSSDNLQLELPELKOKSSNRKRSAGPTVPDRDND---IP--DSLEVEGYVD----- 224
Db 2134 TSSNMEIEKTIYVNSYN---FSDINKINDIDKEKMTLLPMDDELNBCHNIDISLY 2189
Qy 225 ---VKNKRTFLSPWISNIEKKGLT---KYKSSPEKMTAS---DPYSDEKYTGRI 273
Db 2190 NFIIIRIQIKIGNDIKNIKEQENDNICFEYIONNNYFIKSDISIFNKYDHIKVDNYIS 2249
Qy 274 KVSPEARHPVAAAYPIYVD--MENI---ILSKNED-----QSTQ-----NTDSE 314
Db 2250 NNIDVYNNKNSLSSEHVINATNIENIMTSIYEINEDTMSLEETQDKLLEYENFKKE 2309
Qy 315 TTTISKNTSTSR-THTSEVHGNAEVHANTSTSRHTSEVHG-----NAEV 358
Db 2310 KNIIINNRYIVHFNKLEIENSLLEYNSISTNFKINETQNDILKNEFNKIKTKINDKV 2369
Qy 359 -HAVAIDHSLAGERTAEFTWG---LNTADPARL---NANIRYVNTGTADIVVLPPT 410
Db 2370 KELVAVDSTLTLESYOTFNNILYGLDMSNIQDYKKEEDINNVELKVKLYIENITNLGRI 2429
Qy 411 SLVLGNQTLATIKAKENQLS---QILAPNNYPSKNLAPIALNADDESSPTITMANYQ 467
Db 2430 NTFI---KELDKYQDENNGIDKYIEIKENNSYIIR-LKEKANNLEKENS-----K 2476
Qy 468 FLELEKTOQLRDTDOVYGNINATYFENGRAVDTGSMSEVLPQIOETTARIIFNGKDL 527
Db 2477 LQONIKRNETEL-----YNNIN--IKDD-----IMNTGKSV 2505
Qy 528 NLVERRIAIVNSDPLETTKPMTLKELAKIAFGFNEPNCNIOYOGKDTTEDFPNDOOT 587
Db 2506 NNIKOKFSS-----NLPLKEKL---FQMEEMLNT----- 2532
Qy 588 SONIKNOELAEIATNIYVVL-----DKIKLAKKNILIRDKRF-HYDRNNIAVGADES 639
Db 2533 -NNINNETKRISNTDAYNITLQDIENNNKNNNNNINETIDKLIDHIKIHNEKIOAEIL 2591
Qy 640 VYKEAHREYINSTEGLLNDIKDIRKILISGYIVE---IEDTEGLKE--VINDRYD-- 690
Db 2592 IIDDAKRRV---KEITDINIKAFNETENYNNENGVAKSANNIVDEATYLLNELDXF 2646
Qy 691 MLNISLRQDGKTFIDFKKYNDK 713
Db 2647 LKLNELLSHNNN--DIKDLGDE 2667

RESULT 12
Q9BK46
ID Q9BK46 PRELIMINARY: PRT: 3130 AA.
AC Q9BK46:
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE RETICULOCYTE BINDING PROTEIN 2 HOMOLOG A.
OS Plasmodium falciparum.
CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21101060; PubMed=11160005;

RA Triglia T., Thompson J., Caruana S.R., Delorenzi M., Speed T.,
Coman A.F.;
RT "Identification of Proteins from Plasmodium falciparum That Are
RT Homologous to Reticulocyte Binding Proteins in Plasmodium vivax".
RL Infect. Immun. 69:1084-1092(2001).
DR EMBL, AF312916; AAK19244.1; -.
SQ SEQUENCE 3130 AA; 370415 MW; 13D973DB89D82026 CRC64;

Query Match 5 1%; Score 191.5; DB 5; Length 3130;
Best Local Similarity 19.6%; Pred. No. 0.025;
Matches 169; Conservative 134; Mismatches 291; Indels 269; Gaps 41;

Qy 5 KYLIPLMALSTILVSTGNIIEVIOAEVKQENRLNSESSESGILGYFSDINFPQPMV 64
Db 1920 KHVILANSGIIVMSDTTEITPENPLEDNDLN-----LQLYFERKHEI 1965
Qy 65 TSSITGDLSPSELENT--PSENOYFQSAISGFIKVKKSDEYFPATSADNHVTWVVD 122
Db 1966 TSTLENDSDL---ELDHGNSDESIDNLKYVNDIEL-----HMYST---QILKYLDN 2013
Qy 123 -----OEYINKASN-SNKIRLEKGRLOYI-KIO 148
Db 2014 IOKLGDGCDLVKDKELRELSTALYDLKIOITSVINRENDISNNIDISNKLNEIDAIO 2073
Qy 149 YORE-----NPEKGLDFKLYWTDSONKKE 173
Db 2074 YNFEKYKEIFDQVVEEYKTLDDTKNAYIVKKAELIKVNDINKREDIDYFNDLEDEKSL 2133
Qy 174 VSSDNLQLELPELKOKSSNRKRSAGPTVPDRDND---IP--DSLEVEGYVD----- 224
Db 2134 TSSNMEIEKTIYVNSYN---FSDINKINDIDKEKMTLLPMDDELNBCHNIDISLY 2189
Qy 225 ---VKNKRTFLSPWISNIEKKGLT---KYKSSPEKMTAS---DPYSDEKYTGRI 273
Db 2190 NFIIIRIQIKIGNDIKNIKEQENDNICFEYIONNNYFIKSDISIFNKYDHIKVDNYIS 2249
Qy 274 KVSPEARHPVAAAYPIYVD--MENI---ILSKNED-----QSTQ-----NTDSE 314
Db 2250 NNIDVYNNKNSLSSEHVINATNIENIMTSIYEINEDTMSLEETQDKLLEYENFKKE 2309
Qy 315 TTTISKNTSTSR-THTSEVHGNAEVHANTSTSRHTSEVHG-----NAEV 358
Db 2310 KNIIINNRYIVHFNKLEIENSLLEYNSISTNFKINETQNDILKNEFNKIKTKINDKV 2369
Qy 359 -HAVAIDHSLAGERTAEFTWG---LNTADPARL---NANIRYVNTGTADIVVLPPT 410
Db 2370 KELVAVDSTLTLESYOTFNNILYGLDMSNIQDYKKEEDINNVELKVKLYIENITNLGRI 2429
Qy 411 SLVLGNQTLATIKAKENQLS---QILAPNNYPSKNLAPIALNADDESSPTITMANYQ 467
Db 2430 NTFI---KELDKYQDENNGIDKYIEIKENNSYIIR-LKEKANNLEKENS-----K 2476
Qy 468 FLELEKTOQLRDTDOVYGNINATYFENGRAVDTGSMSEVLPQIOETTARIIFNGKDL 527
Db 2477 LQONIKRNETEL-----YNNIN--IKDD-----IMNTGKSV 2505
Qy 528 NLVERRIAIVNSDPLETTKPMTLKELAKIAFGFNEPNCNIOYOGKDTTEDFPNDOOT 587
Db 2506 NNIKOKFSS-----NLPLKEKL---FQMEEMLNT----- 2532
Qy 588 SONIKNOELAEIATNIYVVL-----DKIKLAKKNILIRDKRF-HYDRNNIAVGADES 639
Db 2533 -NNINNETKRISNTDAYNITLQDIENNNKNNNNNINETIDKLIDHIKIHNEKIOAEIL 2591
Qy 640 VYKEAHREYINSTEGLLNDIKDIRKILISGYIVE---IEDTEGLKE--VINDRYD-- 690
Db 2592 IIDDAKRRV---KEITDINIKAFNETENYNNENGVAKSANNIVDEATYLLNELDXF 2646
Qy 691 MLNISLRQDGKTFIDFKKYNDK 713
Db 2647 LKLNELLSHNNN--DIKDLGDE 2667

RESULT 13
ID 09P008 PRELIMINARY: PRT; 4688 AA.
AC 09P008;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE HYPOTHETICAL PROTEIN U0482.
GN U0482.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEVOVAR 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Leikowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
urealyticum.";
RL Nature 407:757-762(2000).
DR EMBL: AE002145; AAF30894.1; -
DR InterPro: IPR001152; Thymosin_b4.
DR InterPro: IPR000626; Ubiquitin.
DR Pfam: PF00240; ubiquitin; 1.
DR SMART: SM00152; THY; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 4688 AA; 534880 MW; B53ABFAFEE1997E CRC64;

Query Match 5.1%; Score 190.5; DB 2; Length 4688;
Best Local Similarity 20.7%; Pred. No. 0.05;
Matches 168; Conservative 133; Mismatches 341; Indels 171; Gaps 39;

QY 18 VSTSGNLEVI-QAEVKQENRLNSESSESSOGLGYYFSDLFQAPMYVTSSTGDLSPS 76
Db 3676 INKTGQKEVLFVSKGKLSNQLYK-----LVQVYILD-NIHNDIDETRIKIFDHWV-S 3725
QY 77 SELENPSENOYFOSAIWSGFIVKKSDEYTFATSDNHVTMVDQEVINKASNSKIR 136
Db 3726 KELEINGVTMISKHGMKSPDTTANFEKIEFQ-----DDNDVLINIDATYKFK 3776
QY 137 LEKGRLYQIKIYORENPTKGLFKLYMDSQKKEVSSDNLQLELQKSSN-----S 192
Db 3777 DEHNINIKQKIVRIKEN-----NDMLIKGI---DNLN-PETYYKLEINIELSK 3820
QY 193 RKRSTSGAPTVPRD-----NDGIPDSLEVEGYTVDKKRTFLS 233
Db 3821 PLKTHNLSASINDKENISLITETGNVLAQVIONDTINDTOQTINVTLSGVNSK-YNG 3879
QY 234 PWISNIHEKKGLTKYSSPEKMWSTADPYSDFEKVTGRIDKNVSPEARHPLVAAYPIVH 293
Db 3880 RQIKVYKKDNNVYESS---LITLQKNDYQLLSNLNSN---REYREFKEIEINHI 3931
QY 294 -----DMF-----NIIISKNEOSTQMTDSE-----TRTISKNTSTSTHSEVHG 335
Db 3932 SNTNNEDELEKLVGSMTFTQTKNTTVQNNDSATIVGRGVNFNKI-KSEDKILENN 3990
QY 336 AEV---HANTSTSRTHSEVHGNAEVAHVALIDHSLSLAGERTMAETGLNT---ADTARL 389
Db 3991 QQVYAFAPKETIRDTNTWLOQYTRPLKDVTSDFK-----EGTNAHDLSSNVNFEETTYK 4045
QY 390 NANIRVYNTGTADIVNLFTTSLVGLKNGQLA-----TKAKENOLSOILAPNVPYPSK 443
Db 4046 LVKIQFNKPTAKKNNINNSNNVILDTNINSINSNYEFTTKVGDHKLINTSSNNVNTNS 4105
QY 444 NLAPIALN-AQDPSSPTITMNY--NQLFELEKTKQLRLTDQVYGYAIATYVFNENGR--- 497
Db 4106 QTINFTLSGVKKSMGKRIKLSKNDTSIHTEWELISNKTQYVILNLNLRKNTYT 4165
QY 498 ---VRVDYGSNMSEVLQIOETTARLIFNGKDLVLERIAAIVNPSDPLETT-----K 547

Db 4166 LIDVKLIDNNNVSDPEKGNLTNSFITRTSAIVNLNIEISNASTLTKSTIIKINLND 4225
QY 548 PDMTLKEA-----LKIAFGNEPENGNIQYQKQITTEDFNFDOOTSNIK-NOLAE 597
Db 4226 PDVNLRRQDQATIVYGNKKQMGFIYVSGNIKYLTAFLVDLNFN-DKVNINISFNKPS 4284
QY 598 LMATINITYVLDKIKLNKAMNLLIDRKRFHYD-----RNNIYAGA---DESVYKAHRE 647
Db 4285 IAAEN-----IGIDKSNNTI-----YNDSTIKLELNNDIYNGPINKETIYKMANOK 4332
QY 648 VINSSTGGLNIDKIDIRKILSGYIVLEIDTEGLEKEVIND---RYDMLNSSL-RQDKTF 704
Db 4333 --NNIDVDLGIQINPKIAHNLR-FLAKFKSTN-----NDIETINVIKSSLVNNDGKTS 4383
QY 705 IDEKKYDKLEPLIYSNPKY-VNYYAYATKENT 735
Db 4384 IRTFLNMLK-----ANKLYSLVDVYVYLVNNS 4410

RESULT 14
ID 09CF64 PRELIMINARY: PRT; 1072 AA.
AC 09CF64;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE UNKNOWN PROTEIN.
GN YQFG.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RA Bolotin A., Wincker P., Mauer S., Jallion O., Malarme K.,
RA Weisenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis.";
RL Genome Res. 0:0-0(2001).
DR EMBL: AE006392; AAK05715.1; -
KW Complete proteome.
SQ SEQUENCE 1072 AA; 113056 MW; 464446E2656CAA08 CRC64;

Query Match 5.0%; Score 187; DB 2; Length 1072;
Best Local Similarity 18.8%; Pred. No. 0.01;
Matches 146; Conservative 151; Mismatches 333; Indels 148; Gaps 29;

QY 7 LIPMLASTIIVSSTGNIQAEVKQENRLNSESSESSOGLGYYFSDLFQAPMYVTS 66
Db 319 LYPISSEASVYDNTLNTLSISLSSISS-----SQTENSQSG-----A 355
QY 67 STTGDLSPSELENPSENOYFOSAIWSGFIVKKSDEYTFATSDNHVTMVDQEV 126
Db 356 STAEIETSDSENSSLSSNO-----INSNSNSEKDSQSSLGSSMSNSESSEHSNSNI 410
QY 127 NKASNSKIR-----LEKGRLYQIKIYORENPTKGLFKL-----YWTDSQNKKEV 177
Db 411 NETNNSSEITNIIPLPSNPTESSNSVSDQTSSEASINSNSISLSPSNISSTSDSASATNS 470
QY 178 DNLQLELKQKSSSKRKRSTSAQYVPRDNDGI-----PDSLEVEGYTVYKKKRTFLS 233
Db 471 DFSVVAEYVANNSLASVNNSSSVLSTSTADNLGINSQSNLTKD--SSEISTSGAFLS 528
QY 234 PWISNIHEKKGLTKYKS---SPEKWS-----TASDPYSDPEKVTGRIDKNVSPEAR 281
Db 529 ---SNQTSSEKSTNSNSISLSPSNISSTVLESTSSNFSNVAEYVANNSLASVNNSS 585
QY 282 HPLVAAYPIVAVDM-----ENIILSKNEQ-----STQNTDSEFTRTISKNTSTSRHTS 330
Db 586 SVLSSTSTADNLEINQGSQDLTKDSSSEISTSGAFLSSNQTSSEASNSMSINSISLST 645

OY 331 EVHGNAEVAHANTSTRTHTSEVHGNAEVAHAVIDHSLSLAGEPTMAETMGLNTADTARLN 390
DB 646 SLTNSSESAITNOSNS-SEATKVDNNSTHSSNLNS-----GSDNSDSDSDSDSDSNL- 699
OY 391 ANIRVYNTGTADIVNLPTTSLVLGKNOTLATIKAKENOLSO-----ILAPNNYPSKNLA 446
DB 700 -----SSSPNLETNOTISSKPSSEVNNIISEPKKVVSSNSNOESTH 741
OY 447 PLALNAODEFSSTPTTMANYNOFLEKTKQLRLDIDQVYGNATYN--FENGRAVVDTG 503
DB 742 EMSTPMKSSISSTSPITSSSQOKESQSN--LINTTEGINNPITFNNSSSENSAAILF- 797
OY 504 SMMSEVLPOIOETARIIF-----NGKDL-----NLVERIAAVNSDPLETTKPKM 550
DB 798 -SYSNNSSSESTGCLYISNEAORDGSEISHSLPSSNSNENNVSSIQOATLESSEKST 856
OY 551 TLK-EALKIAGFNEP-NGNLOYOGKIDTEPFDFNPOQSONIKNOALATNNTIYTD 608
DB 857 MKRSSLSLITNSTSHQNDNOSNSD-----EVKSNNVESILGOLNISKNTHN 908
OY 609 KI--KLAKMNLIRDKRPHYDRNNIAVGADESVYKHAHREYNSSTEGLLNLNDKDIR 665
DB 909 SLTQSLAVIYTLPSKSVTKNEKNSNTVSEKLIKTPQKN-DESONLGQITALDLSFN 967
OY 666 KLSGIYVIEETEGLEKVIN-----DRYDMLNISLRD-GTTFIDFKKYNDKL 714
DB 968 K-----EVTMEDSKTPVDKVLNENGRSONNKTSTIARDKNKVPKRFSEFNSKI 1018

RESULT 15
O46149
ID O46149 PRELIMINARY; PRT: 2178 AA.
AC O46149; O46147; O46148; 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE ALPHA-TOXIN.
OS Clostridium novyi.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC19402;
RX MEDLINE=95342160; PubMed=7616958;
RA Hofmann F., Herrmann A., Habermann E., von Eichel-Streiber C.;
RT "Sequencing and analysis of the gene encoding the alpha-toxin of
RT Clostridium novyi proves its homology to toxins A and B of Clostridium
RT difficile.";
RL Mol. Gen. Genet. 247:670-679(1995).
RN [2]
RP SEQUENCE OF 1204-2178 FROM N.A.
RC STRAIN-ATCC19402;
RA Hofmann F., Habermann E., von Eichel-Streiber C.;
RL Submitted (JUL-1993) to the EMBL/Genbank/DBJ databases.
DR EMBL; Z48636; CAA8565.1; -
DR EMBL; Z23280; CAA80818.1; -
DR EMBL; Z23281; CAA80819.1; -
DR InterPro: IPR002479; CW_binding.
DR Pfam: PF01473; CW_binding.1; 10.
SEQUENCE 2178 AA; 250134 MW; 980ADCE031C4A75A CRC64;

Query Match 4.9%; Score 183.5; DB 2; Length 2178;
Best Local Similarity 20.6%; Pred. 0.043;
Matches 152; Conservative 109; Mismatches 241; Indels 235; Gaps 34;

OY 122 DOEVIKASNSKIRLEKGRLOYI-KIYOYREN-----PTEKGLDFKL---- 163
DB 202 DIIINFLSNYFK-----YDIGKLNQKNNNNKMAIAGATDINTENILTNKLKSYX 253
OY 164 YWTDQONKKEVISSDNLQLELKOKSSNRKRSAGFTVPDRNDGIPDSLE-VEGYT 222

DB 254 YQELIQTNNLAAASDILRLAILKKY-----GYCCIDLDFLPGVN 292
OY 223 VVKNKRFTLSPWISN-----IHEKGLTKYKSSPEKWSADSDYDFEKYTGRI 272
DB 293 LSLFNDISKPNMGDSYWEAIFEAIAINEKRLMNN-----PYKMEQVPSRI 340
OY 273 DKNVSEAR-----HPLVAAPVIVHDMENIILSKNEDOSTONTDETPTISKNTS-TSRT 327
DB 341 KERILSFVNHDINDILPLGDIKISQLEILRLKAATGKRTFSNAFTISNNDLTLNN 400
OY 328 HTSEVHG-----NAEVAHANTSTRTHTSEVHGNAEVAHAVIDHSLSLAGEPTMAETMGL 381
DB 401 LISQLENRYELLNSITIOEKFKICETIYDYSIVNSVELVLETTPKNLMSDGSSEFYQIIGYL 460
OY 382 NTADTARLANIRY-----VNTGTAPIYVNLPTT-SLVGKNOTLATIKAKENOLSOILA 435
DB 461 SSGFKREVSYTFSPSPNITSSATCGTYHFIKNTFPMLSQNDI-----FEA 508
OY 436 PNNYPSKNLAPIALNAODEFSSTPI--TMVNOFLEKTKQLRLDIDQVYGNATYN 492
DB 509 SNMLYFSK-----THDEFKSSWLLRSNIAEKEFOKLITYIGR-----TLN 549
OY 493 FENGR-----VVDGSSNMSEVLPOIOETARIIFNGKDLNVERIAAVNSDPLETTK 547
DB 550 YEDGLNFKWKRYTT-----SELKYLEVNSKTIENYDLNMI-----LQIOG 593
OY 548 PMTLKALKIAGFNEPNGNLOYOGKIDTEPFDFNDO--QTSQNIKNOLAELN----- 599
DB 594 DDISYSAVNV-FGKN-PKKSILIOGVDDFANVFYFENCIYVSDNNILNSRFNDKKIK 651
OY 600 -----ATNIYVLDKIKLN-----AKMNL----- 619
DB 652 LTLIGHGVNFPKLFEGKTVNDLYTNITKPKLOHLREGVILKKKYLIKINILGCMFT 711
OY 620 -----TRDKRFH-----YDRNNIAVGADESVY--KEAHREYNSSTEGLL 658
DB 712 PVDVINSTVKGLEFNKISDLOPKGFSKQOLEISAKTAIRINREKREVLDFGK-WVS 770
OY 659 NIDKDIRKILSGIYV--EIDTEGLEKVINDRYDMLNISLRDGTTFIDFKKYNDKLP 715
DB 771 NMDLIAEQISNKKYVYVWNEVENT-----LSARVEQLN-----KAEFAKQINDSI 815
OY 716 LYISPNYK--VNYYA 729
DB 816 QTTNNOELKOSLVNTYA 832

Search completed: December 2, 2001, 13:55:15
Job time: 550 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:52:01 ; Search time 44.8 seconds
(without alignments)
601.532 Million cell updates/sec

Title: US-09-747-521-4_COPY_1_735
Perfect score: 3765
Sequence: 1 MKRKKVLIPMLSTILVSS.....LYISPNKXVAVYAKENT 735

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3652	97.0	764	1 PAG_BACAN	P13423 bacillus an
2	211.5	5.6	192	1 YPA_BACAN	P13422 bacillus an
3	183	4.9	1803	1 YJL3_YEAST	P47024 saccharomyc
4	177	4.7	1658	1 YM67_YEAST	Q03661 saccharomyc
5	172	4.6	1271	1 Y338_MYCGE	P47880 mycoplasma
6	171	4.5	1230	1 SMC3_YEAST	P47037 saccharomyc
7	166.5	4.4	1928	1 MYS1_YEAST	P08964 saccharomyc
8	160	4.2	1790	1 USO1_YEAST	P25386 saccharomyc
9	159	4.2	2116	1 MYS2_DICDI	P08799 dictyostell
10	158.5	4.2	1487	1 MDS3_YEAST	P53094 saccharomyc
11	158	4.2	1885	1 PAT1_SCHPO	O13735 schizosacch
12	156.5	4.2	803	1 SW16_YEAST	P09959 saccharomyc
13	155.5	4.1	1018	1 FNBA_STRAU	P14738 staphylococ
14	155.5	4.1	1541	1 IGAL_HAEIN	P42782 mycophillus
15	155.5	4.1	1545	1 IG33_HAEIN	P45385 haemophilus
16	155	4.1	1630	1 MSP1_PLAFK	P04932 plasmodium
17	155	4.1	1639	1 MSP1_PLAFM	P04933 plasmodium
18	154.5	4.1	1001	1 S155_YEAST	P43512 saccharomyc
19	154	4.1	1251	1 RBP2_PLAVB	Q00799 plasmodium
20	153	4.1	1570	1 P3K1_DICDI	P54573 dictyostell
21	153	4.1	1744	1 TANA_XENLA	O01550 xenopus lae
22	151.5	4.0	2334	1 MABA_BACSU	O07933 bacillus su
23	151.5	4.0	2869	1 RBP1_PLAVB	Q00798 plasmodium
24	151	4.0	1957	1 YDB6_SCHPO	O10411 schizosacch
25	150.5	4.0	1208	1 PCP1_SCHPO	Q92351 schizosacch
26	150	4.0	1726	1 MSP1_PLAFP	P50495 plasmodium
27	149.5	4.0	1114	1 RHB8_YEAST	O12749 saccharomyc
28	149.5	4.0	1420	1 SRB9_YEAST	P38931 saccharomyc
29	149	4.0	1577	1 HLYA_PROMI	P16466 proteus mir
30	148.5	3.9	979	1 P115_MYCHR	P43508 mycoplasma
31	148.5	3.9	1233	1 YF16_YEAST	P43597 saccharomyc
32	148	3.9	1024	1 RIF3_MOUSE	P97434 mus musculu
33	146.5	3.9	1165	1 YNF4_YEAST	P53950 saccharomyc

34	146	3.9	730	1 GLN3_YEAST	P18494 saccharomyc
35	146	3.9	770	1 ACE2_YEAST	P21192 saccharomyc
36	146	3.9	3418	1 BRC2_HUMAN	P51587 homo sapien
37	145.5	3.9	1093	1 SW14_YEAST	P25302 saccharomyc
38	145.5	3.9	1230	1 ST20_CANAL	Q92212 candida alb
39	145	3.9	635	1 PIB2_YEAST	P53191 saccharomyc
40	145	3.9	976	1 SCPI_HUMAN	O15431 homo sapien
41	145	3.9	1358	1 SIR4_YEAST	P11978 saccharomyc
42	145	3.9	1664	1 INT1_CANAL	P53705 candida alb
43	145	3.9	1726	1 MSP1_PLAFK	P04934 plasmodium
44	144.5	3.8	633	1 IPAA_SHIFT	P18010 shigella fl
45	144.5	3.8	678	1 YNC7_YEAST	P53968 saccharomyc

ALIGNMENTS

RESULT 1	PAG_BACAN	STANDARD;	PRT;	764 AA.
AC	P13423;			
DT	01-JAN-1990 (Rel. 13, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DE	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	PROTECTIVE ANTIGEN PRECURSOR (PA) [CONTAINS: PA-20; PA-63].			
GN	PAG.			
OS	Bacillus anthracis.			
OG	Plasmid PXO1.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1392;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=89172073; PubMed=3148491;			
RA	Welkos S.L., Lowe J.R., Eden-McCutchan F., Vodkin M., Leppla S.H.,			
RA	Schmidt J.J.;			
RT	"Sequence and analysis of the DNA encoding protective antigen of			
RT	Bacillus anthracis.";			
RL	Gene 69:287-300(1988).			
RN	[2]			
RP	DOMAINS.			
RX	MEDLINE=91332080; PubMed=1651334;			
RA	Singh Y., Kimpel K.R., Quinn C.P., Chaudhary V.K., Leppla S.H.;			
RT	"The carboxyl-terminal end of protective antigen is required for			
RT	receptor binding and anthrax toxin activity.";			
RL	J. Biol. Chem. 266:15493-15497(1991).			
RN	[3]			
RP	X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).			
RX	MEDLINE=97192099; PubMed=9039918;			
RA	Petosa C., Collier R.J., Kimpel K.R., Leppla S.H., Liddington R.C.;			
RL	"Crystal structure of the anthrax toxin protective antigen.";			
RL	Nature 385:833-838(1997).			
CC	-1- FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,			
CC	AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE			
CC	DEATH. PA IS THOUGHT TO BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC			
CC	CELLS, THEREBY FACILITATING THE INTERNALIZATION OF LF OR EF. PA			
CC	ASSOCIATED WITH LF CAUSES DEATH WHEN INJECTED, PA ASSOCIATED WITH			
CC	EF PRODUCES EDEMA. PA INDUCES IMMUNITY TO INFECTION WITH ANTHRAX.			
CC	-1- SUBUNIT: ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT PROTEINS, A			
CC	PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN EDEMA FACTOR			
CC	(EF). NONE OF THESE IS TOXIC BY ITSELF. ONCE ACTIVATED, PA FORMS			
CC	HEPTAMERS WHICH INSERT INTO MEMBRANES AND FORM CATION-SELECTIVE			
CC	CHANNELS.			
CC	-1- SUBCELLULAR LOCATION: SECRETED.			
CC	-1- DOMAIN: THE C-TERMINAL PART OF PA IS REQUIRED FOR RECEPTOR BINDING			
CC	AND TOXIC ACTIVITY.			
CC	-1- PTM: PROTEOLYTIC ACTIVATION BY FUZIN CLEAVES THE PROTEIN INTO TWO			
CC	PARTS, PA-20 AND PA-63, THE LATER HEPTAMERIZE.			
CC	-1- SIMILARITY: TO C-PERRINGENS IOTA-B TOXIN AND TO VIP1 TOXINS IN			
CC	BACILLUS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M22589; AAA22637.1; -
DR PDB; IACC; 11-FEB-98.
DR InterPro; IPR003896; Binary_toxB.
KW Toxin; Plasmod; Calcium-binding; Signal; 3D-structure.
FT SIGNAL 1 29
FT CHAIN 30 764 PROTECTIVE ANTIGEN.
FT CHAIN 30 196 PA-20.
FT CHAIN 197 764 PA-63.
FT DOMAIN 30 287 DOMAIN 1, CALCIUM-BINDING.
FT DOMAIN 288 516 DOMAIN 2, HEPTAMERIZATION.
FT DOMAIN 517 764 DOMAIN 3.
FT CA_BIND 206 206
FT CA_BIND 208 208
FT CA_BIND 210 210
FT CA_BIND 217 217
SQ SEQUENCE 764 AA; 85811 MW; 84BB22690FEAAB5 CRC64;

Query Match 97.0%; Score 3652; DB 1; Length 764;
Best Local Similarity 97.6%; Pred. No. 8.7e-175;
Matches 717; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 1 MKRRVILPLMALSTILVSTGNLEVIOAEVQENRLNNESSSGGLGVEFSDLNQQA 60
DB 1 MKRRVILPLMALSTILVSTGNLEVIOAEVQENRLNNESSSGGLGVEFSDLNQQA 60
QY 61 PMVYVSTTGDLSTPSSSELENIPSENOYFQSAIWSGFIKKKSDTYTATSADNVTMAY 120
DB 61 PMVYVSTTGDLSTPSSSELENIPSENOYFQSAIWSGFIKKKSDTYTATSADNVTMAY 120
QY 121 DDQEVYINAKSNKTRLEKGRVYQIKYQRENPTREKGLDFLYTDSQNKKEVYISSDL 180
DB 121 DDQEVYINAKSNKTRLEKGRVYQIKYQRENPTREKGLDFLYTDSQNKKEVYISSDL 180
QY 181 OLPELKOKSSNRKKRSTAGPTVPDRDMDGIPDSLEVEGVVDVKNKRTPLSPMISNH 240
DB 181 OLPELKOKSSNRKKRSTAGPTVPDRDMDGIPDSLEVEGVVDVKNKRTPLSPMISNH 240
QY 241 EKKGLTKYKSSPEKWSIASDPSDEKVTGRIDKNVSPRARHPLVAAPYIVHDMENIL 300
DB 241 EKKGLTKYKSSPEKWSIASDPSDEKVTGRIDKNVSPRARHPLVAAPYIVHDMENIL 300
QY 301 SKNEQOSTONUDSETRTISKNTSTSRTHSEVHGNAEVHANTSTSRTHSEVHGNAEVHA 360
DB 301 SKNEQOSTONUDSETRTISKNTSTSRTHSEVHGNAEVHANTSTSRTHSEVHGNAEVHA 360
QY 361 VAIDHSISLAGERTVAETFGNLTADTARLANIRVYNTGTAFIYVLPPTSLVLKQNTL 420
DB 361 VAIDHSISLAGERTVAETFGNLTADTARLANIRVYNTGTAFIYVLPPTSLVLKQNTL 420
QY 421 ATIKAKENQLSQILAPNNYPSKNLAPIALNAQDDFSSPTITMANYNOFLEKTKQRLD 480
DB 421 ATIKAKENQLSQILAPNNYPSKNLAPIALNAQDDFSSPTITMANYNOFLEKTKQRLD 480
QY 481 TDQVYGNATATYFENGRRVVDGSMWSEYLPQIOETTAIIIFNGDNLIVERRIAAVNS 540
DB 481 TDQVYGNATATYFENGRRVVDGSMWSEYLPQIOETTAIIIFNGDNLIVERRIAAVNS 540
QY 541 DPLETTKPDMTLEKALTAFGFNEPENGNOYQKIDTEDEFNFDOOTSQNIQALAEINA 600
DB 541 DPLETTKPDMTLEKALTAFGFNEPENGNOYQKIDTEDEFNFDOOTSQNIQALAEINA 600
QY 601 TNYIVYLDKIKLNKAMNIIIPDRPHYDRNNIYAGADESVYEAHREVINSSTEGLLNI 660
DB 601 TNYIVYLDKIKLNKAMNIIIPDRPHYDRNNIYAGADESVYEAHREVINSSTEGLLNI 660
QY 661 DKDIRKIISGYIVETEDTEGLEKVINDRYDMLNISSLRQDKTFIDFKKYNKDLPLYISN 720

DB 661 DKDIRKIISGYIVETEDTEGLEKVINDRYDMLNISSLRQDKTFIDFKKYNKDLPLYISN 720
QY 721 PMYKVNVAATKENT 735
DB 721 PMYKVNVAATKENT 735

RESULT 2
YPA_BACAN STANDARD; PRT; 192 AA.

ID YPA_BACAN STANDARD; PRT; 192 AA.
AC P13422;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE HYPOTHEtical 21.6 KDA PROTEIN IN PROTECTIVE ANTIGEN 5'REGION.
OS Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=89172073; PubMed=3148491;
RA Welkos S.L., Lowe J.R., Eden-McCutchan F., Vodkin M., Leppia S.H.,
RT Schmidt J.J.;
RT "Sequence and analysis of the DNA encoding protective antigen of
RT Bacillus anthracis";
RL Gene 69:287-300(1988).

CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M22589; AAA22636.1; -
DR HSSP; P13423; IACC.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 162 182 POTENTIAL.
SQ SEQUENCE 192 AA; 21620 MW; DBC7150AE78F8AFA CRC64;

Query Match 5.6%; Score 211.5; DB 1; Length 192;
Best Local Similarity 34.8%; Pred. No. 0.00019;
Matches 46; Conservative 31; Mismatches 34; Indels 21; Gaps 4;

QY 616 MNILIRDRPHYDRNNIYAGADESVYKHAHREVINSSTEGLLNIDKIRKISGYIEI 675
DB 1 MNILIRDRPHYDRNNIYAGADESVYKHAHREVINSSTEGLLNIDKIRKISGYIEI 675
QY 676 EDTE-----GLKEVINDRYDMLNIISSLRQDKTFIDFKKYNKDLPLYISPN 722
DB 676 EDTE-----GLKEVINDRYDMLNIISSLRQDKTFIDFKKYNKDLPLYISPN 722
QY 723 YKVNVAATKEN 734
DB 723 YKVNVAATKEN 734
QY 113 DDVYVYATKED 124
DB 113 DDVYVYATKED 124

RESULT 3
YXL3_YEAST STANDARD; PRT; 1803 AA.
ID YXL3_YEAST STANDARD; PRT; 1803 AA.
AC P47024; P87192;
DT 01-FEB-1996 (Rel. 33, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSPOSON TY4 207.7 KDA HYPOTHETICAL PROTEIN.
GN TY4B OR YJL113W OR J0780.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;


```
DB 406 SDS-----ESQSAQSESGSDEDFEYKMKNEKSTSEETENISESDQGAADAYTK 456
OY 172 KEVISSDNLQDEL-----KOKSSNRKRRSTACPTVDRDNDGIPDSL 216
DB 457 NVEQGEENDEEPEKDIIRSLQKNFHNKNSSEYSENVLENTDAIYERENQ-IND-- 513
OY 217 EYEGYVAVKNNKTEL---SPWISNHE--KKGITYKSS----PEKSTASDPY---S 263
DB 514 -VEGYDVATGKSVESDHEHSP--DNLGYDLAARALQFOOSRNSNCPOKEOVSESYLGHS 570
OY 264 DEEYVGR-IDKNVSPHAPPLVAAAPYIVHDMENITLKNEDQSQNDQSERITSKNT 322
DB 571 KSNLSGRBLDES---EEOIPLKD-----FTGNNNNLKTDRGDLSSVEI 613
OY 323 STSRHTSEVHGNAAEVH-ANTSTSRHTSEVHGNAAEVAIDHSISLAGEERTWAEITGL 381
DB 614 EYEVKSEKKLDGSTEKEVPLSTDTINNSSLGNEDSIYSLDDAALISENLIDVLMEL 673
OY 382 NTA-----DPAFLMANIRYVNTGAPYIVNLPPTSILVIGKNOGLAT 422
DB 674 KTPPKYEVVISSESVYSTSEYEDNTVAMPPOVEY---TSPFMD-PPNSD---ND---D 721
OY 423 IAKENQSQIILAPNNYYSKNNLAPIALNAQODESSTPITM-----NYNQFLEKKT 474
DB 722 YEKKHDLKSTLAA-----LAPAFYKKDAEFVAGVYKSCLSITSGHTIFHTSKET 773
OY 475 KOLRLDTPDOVGNIAATYFENGFRVYDTSNMSEVLPOIQTETARIIFNGKDLNVERRI 534
DB 774 KQVS-DLDESTEYVPEMENTG---DENKNOSKNFPGVANSDKTEENTD---EKVF 824
OY 535 AAVNSDPLETTKPDMLTKELKIAFGNEPNNGLOYOGKIDTEDPNF-DOQTSON--- 590
DB 825 SAINYTN---VTGSSCCDIETASNVEE---NLRYCEKDMNEAMSSGDCEVKONDG 877
OY 591 IKNOIA-----ELNATNIYVLDKIKLNKAKNNILIRDKRFYDNNNNAVGADESUYKE 643
DB 878 SKTQISFSDSPDNFOESNDNTEFSSTK-----YKRNSDLEDESLEKRE 922
OY 644 AAR-EVIN-----SSTEGLLNIDKIRKILSGYIVETEDTEGLK 682
DB 923 LKRAEYVLDKLEDESEDSYEQDYADPEGNDEGSNENIYKGRKK-----DTLGLV 972
OY 683 EYINDRYDMLNITSLAQDCKTFIDFKKYDKLPLYSNPNKYVNV 727
DB 973 EPENEKVN-----KVHEETLFEANVSSSVNV 999

RESULT 5
Y338_MYCGE STANDARD: PRT: 1271 AA.
ID Y338_MYCGE
AC P47580;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL LIPOPROTEIN MG338 PRECURSOR.
GN MG338.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=756993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Nguyen D.T., Usterback T.R., Saudel D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bolt K.F., Hu P.-C., Lueder T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
[2]
```

```
RP SEQUENCE OF 1023-1114 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bolt K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -I- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
CC ANCHOR (POTENTIAL).
CC -I- SIMILARITY: BELONGS TO THE MG307 / MG309 / MG338 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: U39715; AAC71563.1; -.
DR EMBL: U01809; AAD12341.1; -.
DR TIGR: MG338; -.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Lipoprotein; Membrane; Signal;
KW Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 1271 POTENTIAL.
FT LIPID 27 27 HYPOTHETICAL LIPOPROTEIN MG338.
FT LIPID 27 27 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 1271 AA: 142492 MW; FCE6042067310A70 CRC64;

Query Match 4.6%; Score 172; DB 1; Length 1271;
Best Local Similarity 18.7%; Pred. No. 0.23;
Matches 167; Conservative 133; Mismatches 314; Indels 280; Gaps 38;

OY 14 STLYVSTGNLEYIQ-----AEVKQENRLNE-----SESSSGGLGYVSDLNFQAP 61
DB 317 SKVLISKRNNILSVLKTQVNLAAVVIDQYHLLNNKTELTTTGTGTSNLPDLK 376
OY 62 MYVTS-ITGDISPSELENIPSENOYQSAIMSGEIVKKSDEFTMTSDNHTMNV 120
DB 377 FIKSSATTVMMKSAINTKQOEYISDNNGFN--VKSEFLKINPSLSSGSDNSNTQSPFK 434
OY 121 DQOEVIKANSN--NKIRLEKQ-----RLYOIKYOYRENPTEKGLDFKLY 164
DB 435 QYQALNNSSQTAIFPAVHMESNSQAQVVTNLSLVSLSKTTQKQOQPVYVRGDALY 494
OY 165 -----WTDSONKKEVISSDNLQPELKQKSSNRKRRKSTACPTVDRDNDGIP 213
DB 495 AHHIDGNYFLENSSPNKRNFQAEVLLMRFLQGTNNFSK-----DNVS 540
OY 214 DSLVEGYVAVK--NKRTPLSPMI-----SNTH 240
DB 541 FSYVDLGSNSERSRANRRITTLKLYALTMLENGTSNNGQKQVCDLAKLLKNNITNLS 600
OY 241 E-----KKGITYKSSPERKSTASDPYSEFKYTGRIK-----NVSP 279
DB 601 EPIKKQDFNNSLSQIKSSYESTKAKNLNPRDLANIEKLEQAIYVRANNYIKLKE 660
OY 280 ARHPLV-----AAFPYIVHDMENITLKNEDQSQNT-----DSE 314
DB 661 AKESSIGMCQPLPYKRANGSYPSLAKFPNN-----NSDQSSQTLTLKTTAAITSDNE 715
OY 315 TPTISKNTSTSRHTSEVHGNAAEVAHANTSTSRHTSEVHG-----NAEVHVAIDHSIS 368
DB 716 P-TRENNQTLTLTTEVENKAKEVEKKATYSSQSEITILKLSQDLNDLDDLLIS 774
OY 369 LAGERTWAEITGLNTPADTARLANIRYVNT-----GTAPIVNLPTTSIV 413
DB 775 LITD-----SGRICTVANIKRWYFKNTNSTNNFQDSNKEIKGEFDFDIYQALY 828
OY 414 LKNOGLATIRKKNQSLIAPNNYV--SKNLAPIALNAQODESSPTITMNYNQFL- 469
```

DB 829 IRSMONLT---SKER-----FGYYKDLGSVNSTSTMLOHQAOSHTSSSSVQNTLL 877
QY 470 -----ELKTKQLRLDQVYGNI---AFYFENGSRVRVDGSMNSSEVLPQOETT 517
DB 878 DLAKAFKELDPPNO---DAETKTRFLOALMLLVKNG-----AQNKNLLOQAIPIG 928
QY 518 ARIIFN---GKDLNLYERRIAAVNPSDPLETTKPDWTLKALKIANG-----FNEPNCN 568
DB 929 TRAFVSWYGVYDKN-----PSATVSQKTKSSSSANENFNFELON 969
QY 569 LOY-OGKDIETFPDNFDQOTSONIKQALNATNIYVLDK-----IKLAKMNLIND 622
DB 970 PYVTGSGEIMFN---DKOTPIQPDSL---LESENTYRFTDEFPNNSVALSNKQ---GSSD 1022
QY 623 KRFPHDRNNIAVGADESIVYKAEHREYINSSTEGLLINDKDKRIL-----SGYIYEID 677
DB 1023 KRYFGEFNGLTINSNOS-----ISTASAGLTQOLPNNNGQLITATDKAGALSQYKD 1073
QY 678 TEGLEKVIDN-----RYDMLNITSLRQDKTEYDF---KKY 710
DB 1074 KFTLMSLIKTTSSDALNAGELHRSVAVDNNLSRFSRGEPLISFDNKKKF 1127
RESULT 6
SMC3_YEAST
ID SMC3_YEAST STANDARD: PRT: 1230 AA.
AC P47037:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CHROMOSOME SEGREGATION PROTEIN SMC3 (DA-BOX PROTEIN SMC3).
GN SMC3 OR YJ1074C OR J1049.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H303;
RX MEDLINE=97474309; PubMed=9335333;
RA Michaelis C., Ciosk R., Nasmyth K.;
RT "Cohesins: chromosomal proteins that prevent premature separation of
RT sister chromatids.";
RL Cell 91:35-45(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Rose M., Koetter P., Entian K.D.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA Sor F.J.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC PART OF A CHROMOSOME CONDENSATION MOTOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Y14278; CAA74655.1; -
CC DR EMBL: Z49349; CAA89366.1; -
CC DR EMBL: X88851; CAA61313.1; -
CC DR SGD: S0003610; SMC3.
CC InterPro: IPR003439; ABC_transport.

DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02483; SMC_C_1.
DR Pfam: PF02463; SMC_N_1.
KW Mitosis; ATP-binding; Coiled coil; Nuclear protein.
FT NP_BIND 32 39
FT DOMAIN 172 482
FT DOMAIN 685 1041
FT SEQUENCE 1230 AA; 14136 MW; B1520887780341F CRC64;
Query Match 4.5%; Score 171; DB 1; Length 1230;
Best Local Similarity 20.2%; Pred. No. 0.25;
Matches 174; Conservative 131; Mismatches 335; Indels 220; Gaps 36;
QY 22 GNLEVIQAEVKQPNRLNSESOGGLGYFEDLNFQAPMVY---SSTTGDSITSSSEL 79
DB 201 GNLSKLSSEMEQERKELEKYNELRNKKIYQFLYRELNEVINQMERLDGDY----- 254
QY 80 ENIPSENOYFQ-----SATWSGFIVKKSDEYTFATSNADNHYTMVDDO 123
DB 255 NIYSSSEYIQLDKREDMDIYQSKLSTI-EASLKIKNATDLOQAKRSEISOKLTNY 313
QY 124 EVINK-----ASNSNKRILEKGRLYQIK-IYQRENPTKGLDFKLYWTDSONKKEVIS 176
DB 314 NVKIKDVQOQIESNEQORNLDSATLKEIKSIIEQRQOKSLKILPRYOELT---KEEAMV 369
QY 177 SDNLQLEPELKOKSSNRKKRSTAGPTVPDRDNDGIPDSLEVGYVVDYKAKRTFSPWI 236
DB 370 --KLQASLQKORDILKKGGEYARFKSKDEPDTWIHSEIE-----ELKSS----- 414
QY 237 SNHFE-KKGLTKKSSPEKSTASDPYSDPEKTYGRDKNVSPEARHPLVAAPYIVHVM 295
DB 415 QNLELESQLOMDRTSLRKQYSAID---EELEELDSINGDITGOL-----EDFDS 463
QY 296 ENILSKNEDQSTQNTDSEFRTISKNTSRTTSEYHGNAEVHANTSTRTTSEYHGN 355
DB 464 ELHLKOKLESLDTRKELMRKQKQIQTVELTLDVQNG--QNNVETMRSR-----LAN 517
QY 356 AEVHAVAIDHSLSLAGE-----RTMAETMG-----LNTADTARLNA 351
DB 518 GIINVEITEKLEKISPESEVFGTLGELIKVNDKKTGCEVIGNSLPHIIVDTEETATLM 577
QY 392 NIRYVMTGAPRIYVNLPTSLVGLKNQTLATIAKENQLSQILAPNNVYPSKMLAPIALN 451
DB 578 NELYRKKGGRVTF--IPLNLSLSDVKKPSSNTTQIQFTPLKIKYPERFEKA----- 630
QY 452 AQDDFSPTITMANYNOFLEKTKQLR---LDTFDQV-YGNIAF--YNFENGRVAVDTGSN 505
DB 631 VKHVFQKTIYVKDLGGGLAKKHKLNALITLDDBRADKRGVLGGYLDQHKRTRLESLN 690
QY 506 WSEVLPOIQTARIIFNGDLNLYERRIAAVNPS-----DPLETT 546
DB 691 LNSRSQHKKILBELDFVRNELNDIDTKIDQVGNINIKVSNDRSVYLTNIEVYRTSLNTK 750
QY 547 KPD-MTLKEALK-IAGFNPENGNLYQGGDIPEPDN---EDQOTSQNIKNOLA----- 596
DB 751 KNEKLILLESNLAIILKLELNTNRTFAQERKLTFFENDLLQEPDSSELSEKEKESLTK 810
QY 597 -----ELNATN-----IYTVLDKIKLNKMNLLIDKRFPHDRNNIAVG----- 635
DB 811 EISAANKKLITSDALEGITTTID--SLNAELSKLIPQENDELSKMSVEGDAFIPLQD 868
QY 636 -----ADESVYKAEHREYINSST-----EGILL--N 659
DB 869 ELKELQEKSVGEKQHNVAIVELGTVOREIESIAETNNKKLLEKANNQOORLLKRLDN 928
QY 660 IDDKIRKILSGIYEIEDTGSLKE-----VINRBYDM-----LN--I 694
DB 929 FOKSVETKMTKTTLVYRRELQORIREIGLPEDAVAVNPSDITSDOLLQRLNDMNTET 988
QY 695 SSLRQDCK-TFIDPKKYNDK 713

SEQUENCE 1928 AA; 223634 MW; 6F54C7611F43DC9F CRC64;

Query Match

Best Local Similarity 4.4%; Score 166.5; DB 1; Length 1928;
Matches 162; Conservative 145; Mismatches 291; Indels 247; Gaps 39;

```

OY 20 STGNLEVIQAEVQENRRLNSES-----SSQGLGYYPSDLPFOAPMVVTSSTGDL 73
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 894 TVALENTODLLQKEKNLNKNSLNKRVTSSETLOKOPDDL-----VSEKDEIS 944
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY 74 IPSSLE-ENIPSENOYFQSAIWSGFLTKKKSDEYTFAT--SADNHVTWVDDQEV-INKA 129
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 945 REKLEVAQNLEAHQKIQ-----GLQETIRERATLEKLSKNNELIKQSLDNLCDISKE 999
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY 130 SNS-----NKILE-----KGLYQITQYQRENPTKGLDFKLYWD----- 167
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1000 QSSQSLIKESKLEIKRLKDVINSKEEIKSFNDKLSSESDLDIKLVLEKNCNIA 1059
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY 168 -----SON--KKEVISDN-----LOPELKOKSSNSKKKSTAGPT 203
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1060 MSRLQSLVYENSPLRSKNEFKKKAALNQLNKKSESLKMEKIDNHKKELATES--- 1116
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY 204 VPDRLNDGIPDSLEVEG-YTVQVKNKRTPLSPWISNIHEKKGLTKKYSPEKSTASDPY 262
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1117 -KQRD-----DAVSEHGKITAELKETRI-----QLTEYKSNQK--IKEEY 1154
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY 263 SDEKTYGRIDKVNSEPARPLVAAPYIVYVDMENITLSNEDQSTQNTDSERTSKNT 322
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1155 SNQRFET---KEQEQKRNKSLV-----ESLNDKIKLELA--LSQEI 1192
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY 323 STSRTTSEVHGAEVYHANSSTRTHTS-----EVHGAEVH 359
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1193 SLNQYLKRIKSGNS-VETNISSTRSTSYSDDLKEDIKKYYDQLATLETIRLENE 1251
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY 360 AVAIDHSLSLAGERTAET-MGLNTADTARLANIRVNTGTAPIYVNLPTTSVLKQK 418
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1252 ---IEKKNLISRLRFTETRLASSFEDQKIKQMKKLKLIQDMPSIPDLSTL--NE 1305
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY 419 TLATITAKEKQSLIAPANNIYPSKNIAPIALNAQDDFSPTITMANTQLELEKTKQLR 478
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1306 PLDNCPRKESDINKMLLEVLYL--KROLDETTRAHYD-AENAISSALSHSKFRKITQ--GESS 1360
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY 479 LQTDQVYGNATYVNFENGVRVVDGSMSEVLQIQET---TARIJNGDMLNVERIA 535
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1361 LSSSDYI---KLTFEASEERKVSLEDKLTMPLRDRTNLPVGDIIKRDSTISKEYEIR 1416
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY 536 AVNPSPLETTKPDMLTLEKALITAFGNENPQNLQY-----LKVQDVAKIL 571
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1417 YRK---LENYK---LQELI-----NESNGKLSQTLTLDRQSKSEALISQDLRLQK 1462
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY 572 -----QGDITFEFNFDOOTSQNTKQNL-----AELNATNIYVLDR 609
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1463 DLESTEROKELLSSTIKQOQFENCMDDQGNELRLREHIIHALQOEEDVQKNMASTIEK 1522
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY 610 IKL-NAKMLNLLINDKRFPHVDNRNIAVGADESIVKFAHREYINSTEDELLINIDKDIRKL 668
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1523 LKQONQKREKLWEREM--ERNDSMOLOETLLE-----LKVQDVAKIL 1565
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY 669 SGYIVE-----IEDTEGLEKVEYINDRYDMLNISLRQDKFTIDFKKYNDKLPPLYISNP 721
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1566 SDDLHLKERLHSAVEDRSQYTDIINRLKEELN-CSLKAETFLNKKFATLKVKTLETSTINDS 1624
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY 722 NYKVN 726
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1625 EAKIS 1629
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

RESULT 8

USOL_YEAST STANDARD; PRT; 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)

```

DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTRACELLULAR PROTEIN TRANSPORT PROTEIN USOL1.
GN USOL1 OR INT1 OR YDL058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID:4932.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-X2180-1A.
RX MEDLINE=91185402; Pubmed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.,
RT "A cytoskeleton-related gene, usol1, is required for intracellular
RT protein transport in Saccharomyces cerevisiae."
RL J. Cell Biol. 113:245-260(1991).
RN [2]
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetler M.K., Herman D.J., Bendel C.M., McCellan M., Tao N.,
RA Kendrick K.E.,
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -!- DOMAIN: THE RODLIKE TAIL. SEQUENCE IS HIGHLY REPTITIVE. COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -!- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: X54378; CA38253.1; -
DR EMBL: L03188; AAB00143.1; -
DR EMBL: U53668; AAB6659.1; -
DR PIR: A38455; A38455.
DR HSSP: P80220; 1DIP.
DR SGD: S0002216; USOL1.
DR InterPro: IPR002017; Spectrin.
KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT DOMAIN 1 724 GLOBULAR HEAD.
FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
FT DOMAIN 991 1790 CHARGED (HYPER-HYDROPHILIC).
FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 1172 1786 ASF/GLD-RICH (ACIDIC).
FT CONFLICT 847 847 G->E (IN REF. 2).
FT CONFLICT 924 924 E->K (IN REF. 2).
FT CONFLICT 1253 1253 V->I (IN REF. 2).
FT CONFLICT 1319 1319 I->V (IN REF. 2).
FT CONFLICT 1461 1461 N->S (IN REF. 2).
FT CONFLICT 1461 1461 G->S (IN REF. 2).
FT CONFLICT 1581 1581 I->S (IN REF. 2).
FT CONFLICT 1600 1600 G->V (IN REF. 2).
FT CONFLICT 1661 1661 R->S (IN REF. 2).
FT CONFLICT 1772 1772 D->DEEDDEE (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

```

Query Match

Best Local Similarity 4.2%; Score 160; DB 1; Length 1790;
Matches 142; Conservative 145; Mismatches 288; Indels 206; Gaps 32;

```
QY 23 NLEVIQAEVQKQENRLNLSSESSOGGLGYFSDLNFOAPMVVSTSTGD-----LSI 74
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1023 NIEOLKRTISDLEQKEEIKSSDSSKDEYSOILKEKLETATTADEVNKLISELT 1082
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 75 PSSELE-----NIPSNQYFQSAIWGFIKVKKSDY-----TPAASADNHVTM 118
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1083 TREELAEALAVKNIKNLKLEKLESEKALKVEKNEHLEKEKIQLEKATEYTKQOLNS 1142
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 119 WVDQEVINKASNSKIRLEKGRLYQIKI-----OY-----ORENTEKGL 159
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1143 LRANLESEKHEHDLAQTK-----YEQILANKEROYNEEISOLNDELINSTQOENSIKKK 1199
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 160 DFKLYW-----TDSQKKEVYSSDNQLPELAKOSSNSKRRKSTSGAPYPPDDN 209
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1200 NDELEGEVKAANKSTSEESNKLKSEIDALNLIQIKELKKN-----ETNE 1243
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 210 DGIPDSLE-VEGYTYVDVK---NKRTFLSPWISNIHEKKGLRYKSSPKWSTASDPYSD 265
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1244 ASLEESISVSESYVKIKELDECFNKEVEYSELD-----KLKASEDKNSKYLELOKES 1298
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 266 EKVYGRIDKNVSPEARHPLVAAPYIVHVMENII-LSKNEDOSTQNTDSEFTTKSTNT 324
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1299 EKIKELDAKTE-----LKIQLEKITLNLKAKES-----ESELRLKKTISE 1342
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 325 SRTHTSEVGHNAEVHANTSTRTHTSEVGHNAEVAHAIIDSLAGERTVAETWGLNTA 384
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1343 ERK-----NAEQLKELKNEIQIKNOAFEKERKLLNEG 1375
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 385 DTARLANIRVYNTGTAPIYVNLPTTSLVIGKNQTLATITAKE-----NOLSILAPNNY 439
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1376 SSTIQEISEKINT-----LEDELIRLOENEN-ELKAEIDNTRSELEKYSLSNDE 1424
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 440 YPSKNLAPIALNADDFSS--TPITMANYNOPLLEKTKQLRIIDQVYGNATVNFENGR 497
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1425 LLEEKQNTIK-SLODEILSYMDKITRNDKELLSIRDNKRDLSESLK-----EDLR 1473
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 498 VRVDGSMSEVLPQIOTETRA---ITPNGKDL-----NLVERRIAAVNPSPDLET-TRP 548
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1474 AAQESKAVVEEGLKLEKESSKEKALEKSEKMKLESTIESNETELKSS--METIRKS 1531
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 549 DMTLKEALKIIFGFNEPNGNLOYOGKDIETEDFNPDQOOTSONIKNQL-----AELNAT 601
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1532 DEKLEQSKSA---PEDIKNIQHEKSDLS-RINSEKQIELEKSLKLIKASEGELETV 1587
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 602 --NITYVLKIRLAKNMILIRDKRFHYDRNNIAVGADESVMKAEHREVINSSTEGLLLN 659
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1588 KOELNNAQEKIRINAEETVLKSK-----LEDIERELKMQAE---IK 1627
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 660 IDKDIRKILSGIYVIED-----TEGLKEVINDRKDMILMISSLRQDGKTFIDFKRYN 711
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1628 SNOEKELITSLRLKLEQELDSTQOKAKOSEERRAEVRRKFOVEKSQLDEKAMLETQYV 1687
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 712 D 712
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1688 D 1688
```

```
RT "Conserved protein domains in a myosin heavy chain gene from
RT Dictyostelium discoideum.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RN [2]
RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
RC STRAIN-AX2;
RX MEDLINE=90353583; PubMed=2387408;
RA Lucke-Vielmeier D., Schleicher M., Grabatin B., Wippler J.,
RA Gerisch G.;
RT "Replacement of threonine residues by serine and alanine in a
RT phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
RL FEBS Lett. 269:239-243(1990).
RN [3]
RP PHOSPHORYLATION SITES.
RX MEDLINE=88112226; PubMed=2828113;
RA Wagle G., Noegel A., Scheel J., Gerisch G.;
RT "Phosphorylation of threonine residues on cloned fragments of the
RT Dictyostelium myosin heavy chain.";
RL FEBS Lett. 227:71-75(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX MEDLINE=95345066; PubMed=7619795;
RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutcliffe K., Holden H.M.,
RA Rayment I.;
RT "X-ray structures of the myosin motor domain of Dictyostelium
RT Biochemistry 34:8960-8972(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX MEDLINE=95345067; PubMed=7619796;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
RT truncated head of Dictyostelium discoidium myosin to 2.7-A
RT resolution.";
RL Biochemistry 34:8973-8981(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=96206189; PubMed=8611530;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II).ADP.vanadate complex of the
RT Dictyostelium discoidium myosin motor domain to 1.9-A resolution.";
RL Biochemistry 35:5404-5417(1996).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX MEDLINE=97452580; PubMed=9305951;
RA Guzik A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT "X-ray structures of the MgADP, MgATPgammA, and MgADPPNP complexes
RT of the Dictyostelium discoidium myosin motor domain.";
RL Biochemistry 36:11619-11628(1997).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=98070605; PubMed=9405148;
RA Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;
RT "X-ray crystal structure and solution fluorescence characterization
RT of Mg.2(3')-O-(N-methylanthraniloyl) nucleotides bound to the
RT Dictyostelium discoidium myosin motor domain.";
RL J. Mol. Biol. 274:394-407(1997).
RN [9]
RP FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
RP ACTIVITY THAT IS ACTIVATED BY ACTIN.
RN [10]
RP SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
RP INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
RP LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
RP (MLC-2).
RN [11]
RP SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
RP CORTEX.
RN [12]
RP DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
RP MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN BE FURTHER
RP SPLIT INTO 2 GLOBULAR SUBDOMAINS (S1) AND 1 ROD-SHAPED
RP SUBDOMAIN (S2).
RN [13]
RP DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
RP CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
RP CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
RN [14]
RP PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
```

CC THE ACTIN-ACTIVATED ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2) EDTA ATPASE
 CC ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1
 CC POSITION (688).
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M14628; AAA33227.1; -.
 DR PIR: A26655; A26655.
 DR PIR: S00250; S00250.
 DR PDB: 1MMA: 03-DEC-97.
 DR PDB: 1MMD: 17-AUG-96.
 DR PDB: 1MNG: 03-DEC-97.
 DR PDB: 1MNN: 03-DEC-97.
 DR PDB: 1MND: 17-AUG-96.
 DR PDB: 1MNE: 17-AUG-96.
 DR PDB: 1YOM: 23-DEC-96.
 DR PDB: 1LYK: 28-JAN-98.
 DR Dictydb: D001008; mhca.
 DR InterPro: IPR000048; IQ.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00612; IQ: 1.
 DR Pfam: PF00063; myosin_head: 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR ProDom: PD000355; myosin_head: 1.
 DR SMART: SM00015; IQ: 1.
 DR SMART: SM00242; MYSC: 1.
 DR PROSITE: PS50096; IQ: 1.
 DR Myosin: Coiled coil; Actin-binding; ATP-binding; 3D-structure;
 KM Calmodulin-binding; Methylation; Alkylation; Phosphorylation.
 FT DOMAIN 1 761
 FT DOMAIN 762 791
 FT DOMAIN 817 2116
 FT NE_BIND 179 186
 FT DOMAIN 638 660
 FT DOMAIN 738 752
 FT MOD_RES 130 130
 FT MOD_RES 678 678
 FT MOD_RES 1823 1823
 FT MOD_RES 1833 1833
 FT MOD_RES 2029 2029
 SQ SEQUENCE 2116 AA: 243871 MW: 25C3770BBI5E56A1 CRC64;

Query Match 4.2%; Score 159; DB 1; Length 2116;
 Best Local Similarity 19.2%; Pred. No. 2;
 Matches 154; Conservative 125; Mismatches 291; Indels 232; Gaps 38;

QY 2 KRRKVLPLMALSTIIVSSGNELEVIAEYKQENRLNSESSESSOGLLGYFSOLFQAP 61
 DB 1200 OKKKVELDLEKDSQALEEFAAKQALDKLKKLEQELSEVQTO-----LSBANK- 1249

QY 62 MVTTSSTGOLSPSSLEINPSENQFQSAIWSGFIKVKKSDSEYTRATSDNHTVMVD 121
 DB 1250 NVNSDSTNKLH---ETSFNNIKLELEAEOKA-----KQALEKKRRLGLESELKHVNDLE 1300

QY 122 DQEVINKASNSNKRILEKGRLYQIKIQORENPTKEGIDFLYMTDSONKKE----- 173
 DB 1301 EEKKOKSNEKRYDLER-EYSELKQDLEEVASKAV-----TEAKNKESSELEDEIKR 1353

QY 174 ----VISSDLQLPELKOKSSNSRRKRSTSAPTVPDRDNDGIPDSLEVEGYTVYKNR 229
 DB 1354 QYADVSSRDKSVOLKTLQAKNEELRNTA-----EAAQGLDRAERS-----KKKA 1400

QY 230 TF-LSPWISNHEKKGLTKYKSSPEKKWSTASDPYSDEKVTGRID--KNVSPARHPLVA 286

DB 1401 EFDLEAVAKNLEE---TAKVYAKAKAKAE---TDVYSTSELDADKANSSQ----- 1449
 QY 287 AYPVHVDMENILSKNEDQ-----TQNTDSETRTISKNTSTRTSHSEVHNAEVHAN 341
 DB 1450 ----YVQIKRL-----NEELSELKSVLEADERCSAIAKAKTASLESJKDEIDANN 1500

QY 342 TSTSRTHSEVHNAEVHNAVAIDHSLAGERTYAEITMGTLTADTARLANAIRVNTGTA 401
 DB 1501 AKAAERKSK---ELEVAELESLE-----DKSG-----TVNVEFLRKDA 1540

QY 402 PIYVNLPTSLVKNQTLATIKAKENQLSQLAPNNYPSKNIAPIALNQDDFSSPTPI 461
 DB 1541 EIDDLRAID-----REFESRIKDEDK-----KN----- 1565

QY 462 TMVYVQLELEKTKQLRLTDQVYGNATVYFENGVRVDTGSMVSELPQIQ-ETTARI 520
 DB 1566 --TRKQFADLE-----AKVEAOREVVTID-----RLKKKLESIDIIDLTOLDIETKSR 1613

QY 521 -IFNGK---DLNLYERRIAAVNPDPLETTKPDMLTKALKAIAGFNPNQNLQYOGKDI 576
 DB 1614 KIEKSKKKLEQTLAERRAAEGSSKAAD-----EEIRQVW-----OEV 1652

QY 577 TEFDNFNQDQ-----TSQNIKNQLAELATNITYVLDKIKLNKAMNIIIDKRFHYDRN 630
 DB 1653 DELRAQLDSERAALNASKKIKSLVAF-----VDEVAEQLEDEFLAKDLKLVKAKRA 1703

QY 631 --NIAGADESVYKAEHREVINSTEGILNIDKIRKILSGYVEIE----- 676
 DB 1704 LEVELEEVRODLEEDERSRLEDSKRRLTTEVE-DIKK---KYDAVEQVMTKDEAKK 1759

QY 677 ---DTEGLKEVINDRYDMLNITS-----LRQDKTFT-----DERKYYDK 713
 DB 1760 LTDDVDTLKQLEDEKKLNSEBAKKRLSEENEDFLAKLDAEVYKNSRAEKDRKYEKD 1819

QY 714 LPLVSNPNKYVNYAVTKENP 735
 DB 1820 L-----KDTKYKLNDBAATKTQT 1837

RESULT 10
 MD53_YEAST STANDARD; PRT; 1487 AA.
 ID MD53_YEAST
 AC P53094;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE MD53 PROTEIN (MCK1 DOSAGE SUPPRESSOR 3).
 GN MD53 OR YGL197W OR G1307.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / FY1679;
 RX MEDLINE=97197971; PubMed=9046087;
 RA Coglielina M., Klima R., Bertani I., Delneri D., Zaccaria P.,
 RA Bruschi C.V.;
 RT "Sequencing of a 40.5 kb fragment located on the left arm of
 RT chromosome VII from Saccharomyces cerevisiae.";
 RL Yeast 13:55-64(1997).
 RN [2]
 RP CHARACTERIZATION.
 RA Li M.B., Neugeboren L.;
 RL Unpublished observations (xxx-1997).
 CC -1- FUNCTION: NOT KNOWN; NEGATIVE REGULATOR OF EARLY MEIOTIC GENE
 CC EXPRESSION.
 CC -1- SIMILARITY: TO YEAST YER132C.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X91837; CA62947.1; -
DR EMBL: 272719; CA96909.1; -
DR SGD: S0003165; MDS3.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF01344; Kelch; 2.
DR Melos1s.
KW SEQUENCE 1487 AA; 167073 MW; 768A6FBAB796E447 CRC64;

Query Match 4.2%; Score 158.5; DB 1; Length 1487;
Best Local Similarity 19.1%; Pred. No. 1.3;
Matches 170; Conservative 145; Mismatches 318; Indels 255; Gaps 42;

OY 14 STILVSTGNLEVIQAEKQERLINESSSOGLGYFSDLPNQAQPVNTSS----- 67
DB 593 STILFSPHSSNSSKAVQEGRL-----SSGSLDNF-EKNF--PIFARTVSEAQN 643
OY 68 TTGDSLIPSELENISENOYFOSAIWSGFIKVKKSEYTFATSDNHYTMV----- 120
DB 644 TOPQVAMADAKAPNTPS-----TSDEPSSSSSDKLYSTPHYQRNDEE 686
OY 121 DDOEVIN--KASNSKIRLEKGLYQ--IKIYORENPTKGLDKLYWTD--ONKKE 173
DB 687 DDEDEVSPKPVKSSNSI-----YRPIKTESSSTSSNGMIFRVPKKAATVSNTE 739
OY 174 VISSNLIQLELKOKSSN-----SRKKRSTAGPTVDRNDGIP--- 213
DB 740 ALLESNLSIQETSRSSRSIPSELLRSSISEAHORASHPLTSSPLFSDSGTPCGK 799
OY 214 DSLEVEGYTVDKNKRTELSF-----WISNIHEKGLTKYKSPKMGSTASDP 261
DB 800 QLOQLOQHT--IQNPHNLSPRRFSARSASTSYSSSDRAG-----NSISRST-SDS 851
OY 262 YSDFEKTGRIDKNVSPKARHP----- 283
DB 852 FGT-PPVLGVLVPLPQTRREPNEPPPCPAMSTGNTFRSNTLIDYHMSKASPFSSRR 910
OY 284 -----LVAAPIVHDMENIILSKNEDOSTQVTDSETRISKNTSTSRH 328
DB 911 SSHIGRSSPTETENAFATPPASLDGMLGSLKSGSQTQPRMSFPANETIQTP 970
OY 329 TSEVHGNAEVHANTSTSTHT--SEVHGN--AEVHAVIDHSTSL-----AGEETMAETMGL 381
DB 971 TS-----SNMEMSROSTVSNMDSFDSIQSNFALELEPLTPRSILYMPWPTSTVAFAPETFT 1027
OY 382 NTADTARLANIRIVNTGAPILYVNLPTTSLVLGKNOQLATIKAKENDLSQILLAPNNYYP 441
DB 1028 GOVNSKWLPLPALDLVLVAKIYEIPLLYKLTL--EVLYSILAKKEESLICTSLMET 1084
OY 442 SKNLAPILANQDDFSSTPTW--NYNOFLELEKT-----KLRIDTDOY 484
DB 1085 FRTKILNSYKGEDEKNTYLLSNDYQELKLKVLLENIDNGYDPLDKRSRAQSSST 1144
OY 485 YGNITATYFENGRRVVDGSMWSEVLPOIETTARIIFNG-----KDLNL 529
DB 1145 QESSSSANGE-----KTAAGAGSLF-----SSTNVPTVFAAGGRDHSNGSICGFNSKNI 1196
OY 530 VERRIAAVNPSPDLETTYTPDMTLKALKIATG--FNEPNGMLQYOGADITEFDNFDOQ 586
DB 1197 QGSRSTSGFS-----PRVAKMSLSKEIDPKTFYE-----EYEPKSGKSFDDNDQO 1244
OY 587 TSQNTKN-QLAELN-----ATNIYT-----VLDKILNAKMLLIDKREH 626
DB 1245 TNIGSEFNHLFDNMNGSISSTNSISSDLEKEEQOLODLIERDSAEILLDAR- 1303
OY 627 YDRNNIAVGADESIVKAEHREIVNSSTEGLLINIKDIRKILSGYIVETEDEGLEYIN 686
DB 1304 ---RKK-----EDDKVTKD-----ISNDKRNLYLPHKKNMLKAKEG--KETRDVREEEEERD 1350

OY 687 DRYDMLNLSLRQDKTFIDFKYNDKL-PLYISN--PNYKVNYVAYT 731
DB 1351 FGLGMLSLMKIKIRKAK-HVD--KVDSVDPLRKSAPQSPIRAVGST 1395

RESULT 11
FAT1_SCHPO
ID FAT1_SCHPO STANDARD; PRT: 1385 AA.
AC 013735; 09UT10;
DT 15-JUL-1998 (Rel. 36, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ACTIN INTERACTING PROTEIN 3 HOMOLOG.
GN FAT1 OR SPAC15A10.15 OR SPAC15EL.01.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE OF 1-1033 FROM N.A.
RC STRAIN-972;
RA Murphy L., Harris D., Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1023-1385 FROM N.A.
RC STRAIN-972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Lyne M.H.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP IDENTIFICATION, AND GENE NAME.
RX MEDLINE=20143585; PubMed=10679021;
RA Jin H., Amberg D.C.;
RT "The secretory pathway mediates localization of the cell polarity
RT regulator Alp3p/Budp.";
RL Mol. Biol. Cell 11:647-661(2000).
CC -1- FUNCTION: INVOLVED IN THE ORGANIZATION AND/OR FUNCTION OF THE
CC ACTIN CYTOSKELETON.
CC -1- SIMILARITY: TO YEAST BUD6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z97208; CAB10112.1; -
DR EMBL: AL109770; CAB52420.1; -
KW Coiled coil; Cytoskeleton.
FT DOMAIN 1009 1096 COILED COIL (POTENTIAL).
SQ SEQUENCE 1385 AA; 154325 MW; ABB3D40CCAF7537 CRC64;

Query Match 4.2%; Score 158; DB 1; Length 1385;
Best Local Similarity 18.1%; Pred. No. 1.3;
Matches 141; Conservative 129; Mismatches 282; Indels 228; Gaps 30;

OY 55 DLNQPAPWVYVSTTGDL-----SIPSELENISENOYFOSAIWS-----GFIKVKKS 103
DB 326 DRKFSPKLRPLPSLTKSLDPTPTSLKSPSLKSSPS--SFVQKDVYSRNSNLSRLSQANRS 383
OY 104 DEYFATSDADNHVPMVDDQEVINKAS-----NSN 133
DB 384 NVFPGATD---VTRSVSDHRLISSSTINDGEVAPPLDQSRKTTISSPNSPLSATVLPSTT 439
OY 134 KIRLEKGRLOYIKIOYQSEN-----PTEKGLDFKL-----YMTDSQMKK 172
DB 440 PILLRGSSSTLSVKKQKPMNDGSLTNSPNSIRETEEYASPKLEDJADAEVETDAYSQR 499
OY 173 EVI-----SSDNLQLELKOKSS--NSRKRSTAGPTVDRNDGIP---PDSL 216

```

DB 500 ELLERIOKAESEDETSSEISLOGKLSLPVSTQOEIOPSSVPEASNEIAKEPAVT 559
QY 217 EYEGYIVDVYKNNKRTFLSPWISNHEKKGLTKYKSSPEKMTASDPYSDPEKYGRIIDNV 276
DB 560 ALESTI-----EKKEBAPVISSEKIESGISTISTDTKGLGAPENDSELEBRLLQON- 613
QY 277 SPEARPLVAAYP-----IVHVMENIILSKNEDOSTONTDS 313
DB 614 MAFODEPSYFKHYEYSSSESSSEDEFEKSKOTKGYIISNDSTQVEDESEDSPTMTGA 673
QY 314 EYRTISKNTSTSTRTHTSEVHNAEVHANTSTSTRTHTSEVHNAEVAHAVIDHSIAGER 373
DB 674 SAKLINDPSTIV-----SDVYPKRPASPEVETEPESALVSATSPVTNVPVPEA 725
QY 374 TAAETMGLTADTARLANANIRYNTGTAPYNTLPTSLVGNQTLATIKAKENQLSQI 433
DB 726 VHLST-AFSTAPYSTVSN-----SPLPT----- 749
QY 434 LAPNNYPSKNLAPIALNMODFSTPTTWNYNQLELEKTKOLRLDPTQVGNATYNF 493
DB 750 VAPNPVSGSPSEPT-----SKPEKYPVVSQTEKALPKFLGVDTEKY--FLRYNN 798
QY 494 ENGRVHVDTGSMNSEVLPQIETTARIIFNGK--DLNLVERRIAANVSPDLETTKEDMT 551
DB 799 QTRKAYVESPLSNANELGELFSNVYKISFGSDSYELNIED-----PDTRKISYLEDL 851
QY 552 -LKEAKIAFGFNEPNGNO---YQCKDITEFPNPDQOSTONIKNOALANNTIYT-V 606
DB 852 DLKYSKLSVSEFMFEKODANKREDHSEVSAIOHSSAQNLDHVNNTTHESPSAPTEI 911
QY 607 LDKIKLNAMKNNILIRDKRFHYDRNNIAYGADESVAEAREVINSSTEGGLNLDKIRK 666
DB 912 LEMKA-----IEQNIST-----NHTWDSALASSEHSHK 941
QY 667 ILSGYIV-EIETEGLEKVINDRYDMLNISLRDQKTFI-----DFKK-----YNDK 713
DB 942 LANNFVSPDSIDHKFYQOVYKMQ---LELASTKQISAAPFTRIPDKIDKREINARNEK 998

RESULT 12
SWI6_YEAST
ID SWI6_YEAST STANDARD: PRT: 803 AA.
AC P09959;
DT 01-MAR-1989 (rel. 10, Created)
DT 01-MAR-1989 (rel. 10, Last sequence update)
DT 20-AUG-2001 (rel. 40, Last annotation update)
DE REGULATORY PROTEIN SWI6 (CELL-CYCLE BOX FACTOR, CHAIN SWI6) (TRANS-
  ACTING ACTIVATOR OF HO ENDONUCLEASE GENE) (MBF SUBUNIT P90).
GN SWI6 OR YLR182W OR L9470.8.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
  Saccharomycetales; Saccharomycetaceae; Saccharomycaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
  [1]
RN RP SEQUENCE FROM N.A.
  RA Bredem L., Nasmynh K.;
  RA MEDLINE=88014241; PubMed=28214408;
  RA "Similarity between cell-cycle genes of budding yeast and fission
  yeast and the Notch gene of Drosophila.";
  RL Nature 329:651-654(1987).
  [2]
RN RP SEQUENCE FROM N.A.
  RA STRAIN=S288C / AB972;
  RA Johnson M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
  RA Favello A., Fulton L., Galtung S., Greco T., Kirsten J.,
  RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
  RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
  RA Mardis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,
  RA Rifken L., Riles L., Taich A., Trevasakis E., Vignati D.,
  RA Wilcox L., Woldman P., Vaudin M., Wilson R., Watson R.;
  RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
  [3]

```

```

RP CHARACTERIZATION.
RX MEDLINE=92301535; PubMed=1608451;
RA Dirick L., Moll T., Auer H., Nasmynh K.;
RT "A central role for SWI6 in modulating cell cycle start-specific
RT transcription in yeast.";
RL Nature 357:508-513(1992).
CC - FUNCTION: PART OF A COMPLEX INVOLVED IN CELL-CYCLE-DEPENDENT
CC TRANSCRIPTION. SWI4 AND SWI6 ARE REQUIRED FOR FORMATION OF THE
CC CELL-CYCLE BOX FACTOR-DNA COMPLEX. THE REPEATED ELEMENT IN THE
CC UPSTREAM REGION OF HO (5'-CACGAAA-3') IS CALLED THE CELL CYCLE
CC BOX (CCB).
CC - SUBUNIT: MBF CONTAINS SWI6 AND MBP1. SBF CONTAINS SWI6 AND SWI4.
CC - SIMILARITY: STRONG, TO S.POMBE CDC10.
CC - SIMILARITY: CONTAINS 2 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X06238; CAA29581.1; -
DR EMBL: U17246; AAB67460.1; -
DR PIR: S03161; RCBYWE.
DR TRANSFAC: T00096; -
DR TRANSFAC: T01013; -
DR SGD: S0004172; SWI6.
DR InterPro: IPR002110; ANK.
DR Pfam: PF00023; ank. 2.
DR SMART: SM00248; ANK. 2.
DR PROSITE: PS50088; ANK_REPEAT. 2.
DR PROSITE: PS50297; ANK_REPEAT_REGION. 2.
KW Transcription regulation; DNA-binding; ANK repeat; Repeat.
FT REPEAT 317 346 ANK 1.
FT REPEAT 469 498 ANK 2.
FT DOMAIN 631 640 GLU-RICH (ACIDIC).
SQ SEQUENCE 803 AA; 90559 MW; 9B317FCACEC493C CRC64;

Query Match 4.28; Score 156.5; DB 1; Length 803;
Best Local Similarity 19.14; Pred. No. 0.73;
Matches 156; Conservative 126; Mismatches 296; Indels 237; Gaps 39;

QY 37 LINESSSOGGLGYRSDLFQAPMYVTSSTGDL-----SIPSELENIPENQYFQ 90
DB 20 LTRDSEF-----GHFL-LKHFLPLIQYHDTGNINETNPDSPTDEERN----- 62
QY 91 SAIVSGFIYKKSDSEYFATSDAHVYTMVDDQEVINKASNNKIRLEKGLYQIKIOYQ 150
DB 63 -----KLAHYGIANTDGRGEIWELEKCLQLLNLLNF-----GLFODAEFE 107
QY 151 -----RENPTKGLDFKLYWTSQNKKEYISS---DNLD-----LPRL 185
DB 108 EPTDQDEEDPSHKLLENK--TKSENSKQISSKRINNLDMSLDSDAHRELSPKLK 165
QY 186 K-----OKSSNRKKRSTSGAPTPVPRDNDGIPDSLEVGIVYDVYKNNKRTFL 232
DB 166 KIDTSVIDAESDSTPTNARGKRPNDINKGPS-GDNEENGTDD-----NDRT-A 211
QY 233 SPWISNHEKKGLTK-KYSSPEKMTA--SDPYSDFE---KVTGRIDKNVSPARHPVLA 286
DB 212 GPTTFTHD---LTSDFLSPLKIMKALPSVYVADNEQKKLEAFLORLPE----- 261
QY 287 AYPVHVMENIILSKNEDOSTONTDSETRTISKNTSTRTHTSEVHNAEVAHANTSTSR 346
DB 262 -----IQEMPTSLNNDSSNRNSSEGSNQOQHVSFDLSLQEVN---DAPPTQNL 310
QY 347 THISEVGNAEVH---AVA-----IDHSISLAGERTMAETNG-----LNTADTARL 389
DB 311 NIPVDEGNTPPLHMLTSLANLELVKHSNRLYGDNMGESCLVAVKSVNMYDSQTF 370

```

QY 390 NANIRYV-----NTGAPLYNWLPPTS LV-----LG 415
DB 371 EALLDYVPCLLLEDSSMNTTIIHIIITSGMTGCSAAAKYVLDILMGWIKRONPIQSG 430
QY 416 KQOTLATIKAKENOSQILAPNNYPSKRLAPIALNAODFSSTPTMYNOLFLEKTK 475
DB 431 TNEKSKPRDKNGERKDSILEN--LDLKWITANMLNADQSDNCTCINT----- 476
QY 476 QLRIDTDQYNNIATYNE-----NGRVRVDTGSMSEVLPOIETTARIIFNGSKDLN 528
DB 477 AARLGNISIVDALDYGADPFIAKNSGLRPVDEGAGTS---KLQNT-----NGSDENS 526
QY 529 -LVE-----RRAAVNPSDPLETTKPDMTLKALKALAFENPNQLOYQCK- 574
DB 527 KVVSKGYDQGNKRAKRIKRSOLKNPPTSLINDYQMLN-SISKDYENETVOYNEKL 585
QY 575 DTERDFNPDQOTSONIKQIAEL-NATNIYTVLDKIKLNAKMLIRKRFHYDRNNIA 633
DB 586 EKLHLELNEQREELANSRQILANVQOLKDEYSLMQOLTNLKAGIEEESTREESKKLG 645
QY 634 VGADES-----VKEAHR-EVINSSTEG-LLNIDKDIRRLSGYIEIDTEGL 681
DB 646 IADSSGIDMSSEYDADEPFKEFLSDLEDKLOKNEGDISKLL-----EASKEQI 700
QY 682 KEVINDRDMNLISSLRQDG---KTFIDPKRYNDK 713
DB 701 MEQIRNQLEPAERIQSMLEPPTVLLKARINAYKRNDK 735

RESULT 13

FNBA_STAAU STANDARD: PRT: 1018 AA.

AC P14738;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE FIBRONECTIN-BINDING PROTEIN PRECURSOR (FNBP).
GN FNBA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTC 8325-4;
RX MEDLINE=89098998; PubMed=2521391;
RA Sinaas C., Raucel G., Joensson K., Lindgren P.-E.,
RA Antharamiah G.M., Hoeck M., Lindberg M.:
RT "Nucleotide sequence of the gene for a fibronectin-binding protein
RT from Staphylococcus aureus: use of this peptide sequence in the
RT synthesis of biologically active peptides."
RL Proc. Natl. Acad. Sci. U.S.A. 86:699-703(1989).
RC -1- FUNCTION: THE ABILITY OF BACTERIA TO BIND FIBRONECTIN HAS BEEN
RC PROPOSED AS A VIRULENCE FACTOR ENABLING BACTERIA TO COLONIZE
RC WOUND TISSUES AND BLOOD CLOTS. BINDING OF PLASMA FIBRONECTIN TO
CC THE BACTERIAL SURFACE MIGHT BLOCK ADHESION RECEPTORS ON S.AUREUS,
CC THUS REPRESENTING AN IMPORTANT DEFENSE MECHANISM AGAINST TISSUE
CC INVASION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: J04151: AAA26632.1: -
DR InterPro: IPR001899: Gram_pos_anchor.
DR Pfam: PF00746: Gram_pos_anchor: 1.

DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
KW Signal; Repeat; Cell wall; Transmembrane.
FT SIGNAL 1 36
FT CHAIN 37 1018
FT DOMAIN 37 993
FT TRANSMEM 994 1009
FT DOMAIN 1010 1018
FT REPEAT 545 574
FT REPEAT 575 604
FT DOMAIN 745 878
FT REPEAT 745 782
FT REPEAT 783 820
FT REPEAT 821 859
FT REPEAT 860 878
FT DOMAIN 879 948
FT REPEAT 879 892
FT REPEAT 893 906
FT REPEAT 907 920
FT REPEAT 921 934
FT REPEAT 935 948
FT DOMAIN 982 987
SQ SEQUENCE 1018 AA; 111780 MW; 58175E0020E81F1F CRC64;
CONSERVED IN GRAM-POSITIVE COCCI SURFACE
PROTEINS.

Query Match 4.1%; Score 155.5; DB 1; Length 1018;
Best Local Similarity 19.4%; Pred. NO. 1.1;
Matches 165; Conservative 116; Mismatches 311; Indels 257; Gaps 38;

QY 23 NLEVIOAE-VKQENRLNESSSQGLGYFS-DLNFQAPVAVTSSTGDISL---PS 76
DB 109 NIEVYKEEYVKEAKROYETQOSODNSGDQROVDL---TPKATQNGVAETQVEVQAPR 165
QY 77 SLENIIPSNQYFOGAINSGFIKKKSDPYTATGAD--NHTMVAVDQEVYINKSNSNK 134
DB 166 TASESKPRVYTRNSADVA-----EAKESNAKYETGTDVTSKYVIGSTEGHN--NTNK 216
QY 135 IRLKGRLYQIKIYOQRENPEPKG--LDPKLYWDSQNKKEVISSDNLQPELKOKSSNS 192
DB 217 VERHAGQRAVLKXKLFENGLHQGDYFPDTL---SNVNTFGVSTARKVPEIKKGSVYM 272
QY 193 RKRKSTAGPIVPPRDNDGIPDSLEY-----EGYTVDVKKRPFPLFSWISNIHEK 242
DB 273 ATGEVLEGGKIRYTFETND-IEDKVDVTALEINLFIIDPKTVQFNQNTITSTLNEQTSK 331
QY 243 KDLTRYK-----SPEKWTASDPYSD--PEK-----VTGRIDKNVPE 279
DB 332 ELDDVYKKGIDGNTYANLNGSITETFKANNRFSHVAFIKPNNGKTTSVYVGTLMKGSNON 391
QY 280 ARHPLVAAYPIVHVDMENIISKNDQSTONTSETRTISKNTSTSRHTS---EVHGNA 336
DB 392 GNQPRVRIPE-----YLGNNED-----IAKSVYANTTDITSKFEVTSNM 430
QY 337 EVHANTSTSRHTSEVHGNAEYHAAVLAIDHSLSLAGERTMAETMGANTADARLANI--- 393
DB 431 SCNLMLONNGSYSLNIENLDKYVVHYD---GEY-----LNGTDEVDFTQWCH 477
QY 394 -----RVVNGTATIVAVLPPTSVLKQNTLATIKAKENOSQILAPNNYF----- 441
DB 478 PEOLKYIYDRS---YLTWDNGLALYSN-----KANGKNKPITQNNFEYKEKDTI 527
QY 442 -----SKNLAPIALNAODFSSTPTMYNOLFLE-----LEKTKQLRLD- 480
DB 528 KETLNGQYDKNL--VTYVEEYDSSTIDIDHTAIDGGGVDGYIETIETDSALDI 584
QY 481 -----TDQYNNIATY-----NFEEN-----GRVAVDTGS 504
DB 585 DYHTAVDSBAGHVGCTESSESNPIDFEESTYHNSKHADVVEEDTNPGGQVTTES 644
QY 505 NMSEVLPQIQTETARIINGKQKLNLYERRIAVAVNSDPLETTKPRMTLKEALKAFNGNE 564
DB 645 N-----LVEPDESTKGIVTG-----AVSDHTTVEDTKKEYTTESNLIELVDLPE 689

```

QY 565 PNGNLGYGKDIETEFNFDOQTSONIKNOLAELNATNITVLDKIKLNKML----- 619
D 690 EHQAGQGVPEITKNNHHSH-----SGLGTENGNGNDVIEIEIENSHVDKSLGY 742
QY 620 -----IDKRFHYDRNNIAGADESVYEAH-----REVINSSTE----- 654
D 743 EGGQNSGNSGFEDTEEDKPKYEGGNGNIDDFDSVPQIHGQNGKNSGFEDTEEDKPKY 802
QY 655 -----GLLNTDKDRIKILSG-----IYEDTEGKLEKVIDRMDLNISSLRDQKFTFD 706
D 803 EHGNGNIIDDFDSVPQHIFGNKHTTEITE-EDTKDKR-----PSYQFGHNSVD 849
QY 707 FKKYNDKLP 715
D 850 FEE--DTLP 856

RESULT 14
IGA1_HAEIN
ID IGA1_HAEIN STANDARD; PRT; 1541 AA.
AC P42782;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE IMMUNOGLOBULIN A1 PROTEASE PRECURSOR (EC 3.4.21.72) (IGA1 PROTEASE).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX Haemophilus.
RN NCBI_TaxID=727;
RP SEQUENCE FROM N.A.
RC STRAIN-HK368 / SEROTYPE B;
RX MEDLINE=89379374; PubMed=2506130;
RA Poulsen K., Brandt J., Hjorth J.P., Thøgersen H.C., Killian M.;
RT Cloning and sequencing of the immunoglobulin A1 protease gene (iga)
RT of Haemophilus influenzae serotype b.;
RL Infect. Immun. 57:3097-3105(1989).
RN [2]
RP MUTAGENESIS OF SER-288.
RC STRAIN-HK368 / SEROTYPE B;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Killian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases.";
RL J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: CLEAVAGE OF IMMUNOGLOBULIN A MOLECULES AT
CC CERTAIN PRO-1-XAA BONDS IN THE HINGE REGION. NO SMALL MOLECULE
CC SUBSTRATES ARE KNOWN.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X64357; CAA45708.1; -
DR EMBL; M87492; AAA24969.1; -
DR MEROPS; S06.001; -
DR InterPro; IPR000710; IGA_S6.
DR Pfam; PF02395; IGA1; 1.

```

```

DR PRINTS; PR00921; IGASERPTASE.
KM Hydroxylase; Serine protease; Transmembrane; zymogen; signal.
FT SIGNAL 1 25
FT CHAIN 26 1008
FT PROPEP 1009 1541
FT ACT_SITE 288 288
FT MUTAGEN 288 288
SO SEQUENCE 1541 AA; 169370 MW; CE7257CB3196C600 CRC64;

Query Match 4.18; Score 155.5; DB 1; Length 1541;
Best Local Similarity 20.18; Pred. No. 2;
Matches 129; Conservative 98; Mismatches 225; Indels 191; Gaps 33;

QY 21 TGNLEVIOAEVYKQENRLNSESSESS-----QGILG-----YFSDL-NPQA-DMVY 64
D 857 TGNSDVHQLDLANGHINHNSADNSNNVTKYTLTVNSLSGSGSFYVLDLSNKGDKVYV 916
QY 65 TSSYTGDLSPSELENIPESENQYFQSAIWSGFIYKKSDEYTPATSDNHTM--WVDD 122
D 917 TKSATGNFTLQVADKTGEPNINEL-----TLFDASKAQDRHLNVLVNGTVLDGAW--- 967
QY 123 QEVINKASNSKIRLEKGRLYQIKYQRENPTKGLDFKLYWTDSONKKEVSSDLQL 182
D 968 -----KYLKNVNGRDLYNPEVEKRNQT-----VDTN---ITPNNIQA 1005
QY 183 PELKQKSSNSKKRSTSA-----GTYVDRDNDGIPDSLEVEGYTVYKKNRTLSFWS 237
D 1006 DVPSVPSNNEIADVDAVPPAPAPPTSETTEVVAENSKESTVE-KNQ----- 1056
QY 238 NIHEKKGLTKYKSPKWKSTADSPYSDPEKVTGRIKNNVSPKAPRPLVAAVPIYVDMEN 297
D 1057 -----DATETTAO-NREVAEAKSNVKA-----NTQTNE 1084
QY 298 IILSKNEDOSTONTDS-ETRTISKNSTSRTHTESEVGNMAEVHANTSTSRTHTESEVHNA 356
D 1085 VAQSGSEFKETQJTEKETATVEKE-EKAKVETKTEQVFPVTSQVSKQOSSEFVPOA 1143
QY 357 EVHVAVALDHSISLAGERKMAETGLNLTADTAR-----LNANIRY-----VNTGTAPLYN 405
D 1144 E-PARENDPTVNIKEPOSQNT---TADTEQPAKETSSNVEQPVTESTVYNTNSVYEN 1198
QY 406 VLPPTSLVLGKNGQTLATIKAKENQLSLAPNNYPS-----KMLAPALNAOD----- 454
D 1199 -----PENTTPATITQPVTVNSSESKPKKRRRSRVSYPHVAVEPATITSSNRSTVAL 1249
QY 455 -DFSSTPTTMNYN-----QFLELEKTKQLRLDTQY-YGNIAITYNEENGRVAVDT 502
D 1250 CDLST---NTNVLSDARAKQFVALNKGAVSOHSIQLENNNEGQYV----- 1296
QY 503 GSNSEVLPQIQETTARIIFNGKDLNVERIAAVNPSDPLETTPKPMTEKLEAKIAFGF 562
D 1297 ---W-----VSNTSMKNKNVSSQY---RRFSKSTQYQGW---DOTISNNYQLGQVF 1340
QY 563 N-ENGNLQYQKDIETEFNFDOQTSONIKNOLAELNATNITV 604
D 1341 TYVRNSN-----NEDKATG---KNTLAQVNFYSKY 1367

RESULT 15
IGA3_HAEIN
ID IGA3_HAEIN STANDARD; PRT; 1545 AA.
AC P45385;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE IMMUNOGLOBULIN A1 PROTEASE PRECURSOR (EC 3.4.21.72) (IGA1 PROTEASE).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX Haemophilus.
RN NCBI_TaxID=727;
RN [1]

```


RP SEQUENCE FROM N.A.
 RC STRAIN-HK393 / NCTC 8467 / SEROTYPE B;
 RX MEDLINE-92234949; PubMed-1373717;
 RA Poulsen K., Reinholdt J., Killian M.;
 RT "A comparative genetic study of serologically distinct Haemophilus
 influenzae type 1 immunoglobulin A1 proteases.";
 RL Bacteriol. 174:2913-2921(1992).
 CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
 CC PRODUCING INTACT FC AND FAB FRAGMENTS.
 CC -1- CATALYTIC ACTIVITY: CLEAVAGE OF IMMUNOGLOBULIN A MOLECULES AT
 CC CERTAIN PRO-1-XNA BONDS IN THE HINGE REGION. NO SMALL MOLECULE
 CC SUBSTRATES ARE KNOWN.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
 CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
 CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
 CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M87490; AAA24967.1; .
 DR MEROPS: S06.001; .
 DR InterPro: IPR000710; IGA_S6.
 DR Pfam: PF02395; IGA1; 1.
 DR PRINTS: PR00921; IGASERPTASE.
 KM Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 1012 IMMUNOGLOBULIN A1 PROTEASE.
 FT PROPEP 1013 1545 HELPER PEPTIDE (POTENTIAL).
 FT ACT_SITE 292 292 PROBABLE.
 SQ SEQUENCE 1545 AA; 170627 MW; 3EDD75398BF6D478 CRC64;

Query Match 4.1%; Score 155.5; DB 1; Length 1545;
 Best Local Similarity 20.1%; Pred. No. 2;
 Matches 129; Conservative 98; Mismatches 225; Indels 191; Gaps 33;

QY 21 TGNLEVYQAEVROENRLNESESSS-----QGLG-----YFSDL-NFOA-PMVY 64
 DB 861 TNSDVHOLDLNGHILNLSADNSNNVTKYNTLYVSLSGNSFYLLTDLNKGDKVYV 920
 QY 65 TSSGTGDLSPSELENTISENOYFOSAIWISGFIKVKSDSEYTFATSDNHVTM--WVDD 122
 DB 921 TKSATGNFTLOYADKTGEPNHNL-----TLFDASKAQRDHLNLSLVGNFTVDLGAM--- 971
 QY 123 QEVINKASNSKIRLEKGRLYQIKIORYENPTEKGLDKLWTSQNKKEVISSDNIOL 182
 DB 972 -----KYLRLNNGRYDLNPEVEKRNQ-----VDTTN--ITTPNNITQA 1009
 QY 183 PELKQSSNSRRKSTSA-----GPTVPDRDNDGIPDSLEVGTYVDVKNKRTFLSPWIS 237
 DB 1010 DVPSPVSNNEELARVDEAVPPAPATPSETTETVAENSKQESKTYVE-KNEQ----- 1060
 QY 238 NIHEKKGLTKYSSPEKWSADPYSDFEKVGRIDKNVSPEARHPLVAAYPIVHVDMEN 297
 DB 1061 -----DATETTAQ--NREYAKKAKSNVKA-----NTQTNE 1088
 QY 298 IILSKNEDOSTQNTDS-ETRTSKNTSTSRHTSEVHGNAEVAHNSTSRHTSEVHGNA 356
 DB 1089 VAQSGSETKETQTTETKETATYEKE-EKAKVETKTEQVPKVTSOVSPOKQOSEYVPOA 1147
 QY 357 EVHAYVAIDHSLSLAGERTWAEFTMGTLADTAR-----LNANIRY-----VNTGTAPIYN 405
 DB 1148 E-PARENPTVVIKKEQSGTNT-----TADTQPAKETSSNVQEPYESTTVNTGNSVVEN 1202

QY 406 VLPTTSVLGKQNTLATIKAKENQSLIAPNNYPS-----KNLAPIALNAD----- 454
 DB 1203 -----PENTPATPTQPTVNSESSNPKRNRHSVSVSHVNEPATSSNDSTVAL 1253
 QY 455 -DESSPTPTMNTN-----QFLEKTYQLRIDQV-YGNATAYNFENGRRVDT 502
 DB 1254 CDLTST-----NTNAVLSDARAKAQFVALVWGKAVSHISQLEMNNEGQYV----- 1300
 QY 503 GSNWSEVLQIQIETTRILFNKRDNLVERRIAAVNPSDPLETKRDMTLKELAKIARGF 562
 DB 1301 ---W-----VSNTSKNNKYSSQY---RRFSSTQTOGLW---DOTISNNVQLGVF 1344
 QY 563 N-EPNGNLQYQKIDTEFDNFDOOTSONIKQIAELNATNY 604
 DB 1345 TYVRNSN-----NFDKATS---KNTLQAVNIFYSKY 1371

Search completed: December 2, 2001, 13:52:29
 Job time: 384 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:51:18 ; Search time 72.22 Seconds

(without alignments)
775.246 Million cell updates/sec

Title: US-09-747-521-4_COPY_1_735

Perfect score: 3765

Sequence: 1 MKRRKVLPLMALSTILVSS.....LYISNPNTKVVAVTKENT 735

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 segs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR68:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3652	97.0	764	2	I39934 protective antigen
2	877.5	23.3	875	2	I40862 Iota toxin compone
3	211.5	5.6	192	2	I39933 cryptic protein -
4	211.5	5.6	204	2	G59104 hypothetical prote
5	190.5	5.1	4688	2	F82885 hypothetical prote
6	187	5.0	1072	2	A86827 hypothetical prote
7	183.5	4.9	2178	2	S55605 alpha-toxin - Clos
8	183	4.9	1465	2	S31262 TYB protein - yeas
9	183	4.9	1803	2	S56894 TYB protein - yeas
10	179.5	4.8	1051	2	T18351 lmp1 protein - Myc
11	179.5	4.8	1365	2	T30822 lmp1 protein - Myc
12	177.5	4.7	1467	2	PC1253 TYB protein - yeas
13	177	4.7	1658	2	S55101 hypothetical prote
14	176	4.7	1193	2	S62818 botulinum neurotox
15	175	4.6	1802	2	S52611 TYB protein - yeas
16	174.5	4.6	4919	2	T31105 hypothetical prote
17	173.5	4.6	2269	2	T28677 rhoptry protein -
18	173	4.6	1302	1	JC6009 surface-located me
19	172	4.6	1271	2	D64237 hypothetical prote
20	172	4.6	1939	2	T18372 repeat organellar
21	171.5	4.6	2529	2	B64635 toxin-like outer m
22	171	4.5	1230	2	S56850 SMCI protein homol
23	170.5	4.5	1837	2	T41023 probable nuclear p
24	170	4.5	1315	2	T28679 fibrinogen-binding
25	169.5	4.5	1620	2	S61535 nucleotide-binding
26	168	4.5	2523	2	T18477 hypothetical prote
27	168	4.5	3724	2	T18427 hypothetical prote
28	166.5	4.4	1115	2	T41342 probable collid-co
29	166.5	4.4	1230	2	E64664 outer membrane pro

30	166.5	4.4	1928	2	S46773 myosin heavy chain
31	166.5	4.4	2401	2	T28676 rhoptry protein -
32	166.5	4.4	2526	2	T20531 hypothetical prote
33	166	4.4	821	2	S67087 hypothetical prote
34	164.5	4.4	1553	2	T18502 hypothetical prote
35	163.5	4.3	1014	2	H71602 protein with DnaJ
36	161.5	4.3	1595	2	T31082 endo-1,4-beta-xyla
37	161.5	4.3	4152	2	T31102 filamentous hemagg
38	161	4.3	1538	2	T29095 cardiac muscle fac
39	161	4.3	2399	2	H71879 toxin-like outer m
40	160.5	4.3	888	2	A38539 p101 protein precu
41	160.5	4.3	1979	2	C71622 hypothetical prote
42	160.5	4.3	2340	2	B71704 cell surface anti
43	160	4.2	1790	2	S67593 transport protein
44	159	4.2	1516	2	E71619 RAD2 endonuclease
45	159	4.2	2116	2	A26655 myosin heavy chain

ALIGNMENTS

```
RESULT 1
I39934
C:Species: Bacillus anthracis
C:Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 01-Dec-2000
C:Accession: I39934; S69160; F59104
R:Melkos, S.L.; Lowe, J.R.; Eden-McCutchan, F.; Vodkin, M.; Leppla, S.H.; Schmidt, J.
Gene 69, 287-300, 1988
A:Title: Sequence and analysis of the DNA encoding protective antigen of Bacillus ant
A:Reference number: I39933; MUID:89172073
A:Accession: I39934
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-764 <RES>
A:Cross-references: GB:M2589; NID:g143280; PIDN:AAA22637.1; PID:g143282
R:Fieldman, T.C.; Gordon, V.M.; Leppla, S.H.; Kimpel, K.R.; Birch, N.P.; Lon, Y.P.
Arch. Biochem. Biophys. 316, 5-13, 1995
A:Title: In vitro processing of anthrax toxin protective antigen by recombinant PC1 (
A:Reference number: S69160; MUID:95142670
A:Accession: S69160
A:Molecule type: protein
A:Residues: 197-202 <FR1>
R:Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koeh
J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of pXOI, the large Bacillus anthracis plasmid harb
A:Reference number: A59091; MUID:99445483
A:Accession: F59104
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-313, 'Q', 315-764 <OKI>
A:Cross-references: GB:AF065404; NID:g4894216; PIDN:AAD32414.1; PID:g4894326
A:Experimental source: strain Sterne
A>Note: similar to anthrax toxin moiety, protective antigen, pagA formerly pag, plasm
C:Genetics:
A:Gene: pXOI-110
A:Genome: plasmid
C:Function:
A:Description: three component exotoxin; protective antigen binds to receptors on the
Y active components edema factor or lethal factor; the complex is internalized by rec
F:1-29/Domain: signal sequence #status predicted <IG>
F:30-196/Domain: propeptide #status predicted <PRO>
F:197-202/Product: protective antigen #status experimental <MAT>
```

```
Query Match 97.0% Score 3652: DB 2: Length 764:
Best Local Similarity 97.6% Pred. No. 8e-182:
Matches 717: Conservative 2; Mismatches 16; Indels 0; Gaps 0;
OY 1 MKRRKVLPLMALSTILVSSGTGNLEVIOAEVKQENRLNSESSESSQGLGYFSDINQOA 60
DB 1 MKRRKVLPLMALSTILVSSGTGNLEVIOAEVKQENRLNSESSESSQGLGYFSDINQOA 60
```

Qy	61	PMVYTSSTTGDLSTPSSLELEINPSENOYFQSAIWSGFIKVKSDERTPATSAHNHTMAY	120
Dp	61	PMVYTSSTTGDLSTPSSLELEINPSENOYFQSAIWSGFIKVKSDERTPATSAHNHTMAY	120
Qy	121	DDQEVYKAKSNNKIRLEKGLVQIKIQYORENPTEKGLFKLYWDSQNRKEVSSDNL	180
Dp	121	DDQEVYKAKSNNKIRLEKGLVQIKIQYORENPTEKGLFKLYWDSQNRKEVSSDNL	180
Qy	181	QLEPCLKOKSSNRKRKSTASAGPYVPBDNDNGIPDSLEVEGYTVDVANKRFLSPWISNH	240
Dp	181	QLEPCLKOKSSNRKRKSTASAGPYVPBDNDNGIPDSLEVEGYTVDVANKRFLSPWISNH	240
Qy	241	EKKGLTKYKSSPEKWSASTADPYSDFEKYVTRIDKNVSPEARHPLVAAYPVIVHDMENIL	300
Dp	241	EKKGLTKYKSSPEKWSASTADPYSDFEKYVTRIDKNVSPEARHPLVAAYPVIVHDMENIL	300
Qy	301	SKNEDQSTQNTDSTRTISKRTSTSRTHSEVHGNAEYHANISTSKRTHSEVHGNAEYHA	360
Dp	301	SKNEDQSTQNTDSTRTISKRTSTSRTHSEVHGNAEYHASFDDIGGSAGFSNNSST	360
Qy	361	VAIDHSLSLAGERIWAETMGLNTADTARLANANRYVNTGAPLYNYLPTTSLVLGKNOL	420
Dp	361	VAIDHSLSLAGERIWAETMGLNTADTARLANANRYVNTGAPLYNYLPTTSLVLGKNOL	420
Qy	421	ATIKAKENQUSOLIAPNNYYPKSNLAPIALNAODFSPETITWNYQPLELEKTKOLRD	480
Dp	421	ATIKAKENQUSOLIAPNNYYPKSNLAPIALNAODFSPETITWNYQPLELEKTKOLRD	480
Qy	481	TDQYVGNATYVNFENGVRADVDTGSNMSEVYPOIOWETARIIFNGKOLNVERRIAAVNS	540
Dp	481	TDQYVGNATYVNFENGVRADVDTGSNMSEVYPOIOWETARIIFNGKOLNVERRIAAVNS	540
Qy	541	DPLETTKPDMLKEALKIAEGFNEPNQNLQYOKKDITTEPFENFDQOTSONIKNOJLAELNA	600
Dp	541	DPLETTKPDMLKEALKIAEGFNEPNQNLQYOKKDITTEPFENFDQOTSONIKNOJLAELNA	600
Qy	601	TNITVTVDIKIRLANKNMILIRDKRFPHDRNNIAVGADESVEYKAAHEVINSSTEGLLNI	660
Dp	601	TNITVTVDIKIRLANKNMILIRDKRFPHDRNNIAVGADESVEYKAAHEVINSSTEGLLNI	660
Qy	661	DKDRIKILISGYIVEIEDTEGLEKVEINDRYMNLISSLRODGKFIIDFKKYNKRLPLYISN	720
Dp	661	DKDRIKILISGYIVEIEDTEGLEKVEINDRYMNLISSLRODGKFIIDFKKYNKRLPLYISN	720
Qy	721	PNYKVNVNYAATKENT 735	
Dp	721	PNYKVNVNYAATKENT 735	

[illegible]

Db	15	TAMISOTLSYVWVAQTTQNDTQNDKKEITNENTLSSNGLGYEAFDEHFKDLELMAPIKN	74
Oy	70	GDLSPSELENIJSE-NQYFQSAIWSGFIKVKKSDDEYATATASADNHVYTMWDDGEVUK	128
Db	75	GDLPEEKKYDKDLLENTSSIKISIRMTGRITIPSEDDGYILSTOR-MDVLMOINAKDIAIK	133
Oy	129	ASNSNKIRLEKGRLYQIKIJOYRENN-----PRXKGLDFKLYMTDSOONKEVSSPN	179
Db	134	---TLKVMKKGQVXNRIEILQDKNLGSDIDLNSVP-----KLYM-ELUNGKTYIPEEN	182
Oy	180	LQDELKOKSSNKKRSTASGPTVP-----DRDNDGJPDSDLEVEG	220
Db	183	LEFRDYSKIDEND-----PEIPANNFEVDFESAAMEDEDLTDNDNDIPDAEYKNG	233
Oy	221	YTVQVKKRFTLSWISNIHEHKKGLTKYKSSPEKMSIASPSYSDFEKVIYGRIDKANVSPA	280
Db	234	YTI---KDSIAVKNWDSFAE-QGKRYKSYSLIESNAGBPYVIDYKASGSDIKALIKTEA	288
Oy	281	RHPLVAAPYIVHVDENIILSKNEDOSTQWTDSEFRTIKSTSTSRHTHSEVHNAEVA	340
Db	289	RDLPLVAAPYVGVOMENLIISTNEHSS-----DQKTVSATINSTKDANTV--GVYSIA	342
Oy	341	-----NTSFTHTSEVHNAEVAHAVIDHSLSLAGERWAEETMGANTADTARLAN	392
Db	343	GYONGFTGNITTSYSHTTD-----NSTAVODS---NGE-SWNTGLSINKGESAVINN	391
Oy	393	IRRYNTGTAPRYANLPPTSLVLGKNQGLATIRAKKENQLOSLAPNNYYSKULAPALNA	452
Db	392	VRYNTGTAPRYAKYKTPPTNLVL--DEGLATIRAKADONIGNISPNTYPRKGLSPALMT	450
Oy	453	QDDSPSTPTMNYNOPELEKTEKQOLRIDTQOYVGNIAIFYENGRRVYDGSMSWEVLPQ	512
Db	451	MDQGNALLPIINDQLKLDKSDGOKIKETIYQVSGNVTGRK-SQGOI-ITEGSMWSYISQ	508
Oy	513	IQETTAIIFNGKDLNVERRIAANVESDPLETTKQDMTLKEALKITAFGEPNPNGLOY-	571
Db	509	IDSVASAIILID-TGSOTFFERRAVAKEGCNPEDKTP-PEITIGEAIRKAFSATK-NGELIYF	565
Oy	572	QGRKITE--EDFNPDQOSTONIKNOLAELNATNYIVLDKIKYKAMNMLIINDKRP--HY	627
Db	566	NGPIDESCVELLIPDNTSLTIKEQLYIDDKKIYV---KLERGMNLIKVPSTYFNE	621
Oy	628	DR--NNIIVGADSEVYKAEHREVINSSTEGE-----LNDIKDIRKILSGY-----	671
Db	622	DEVNMP--ASMSINDIKQDDGLOSVANKLSGRTKILIMSKLAPYKRVFESGYKDPST	679
Oy	672	---IYEDTEGLKEYINDRYMLNIS-----SLRQDGKTFPIDEFKRYND	712
Db	680	SNSITIVIKRKEOKTDVLVEKDYTFKSYEFETTGKDDSDIETITLSSGVIFLDMNLSTE	739
Oy	713	--KLPILISNPYK 725	
Db	740	LNSTPEILKEPEIKV 754	

Query Match	5.6%	Score 211.5	DB 2	Length 192
RESULT 3				
I39933				
cryptic protein - Bacillus anthracis				
C:Species: Bacillus anthracis				
C:Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 15-Oct-1999				
C:Accession: I39933				
B:Winkos, S.L.; Lowe, J.R.; Eden-McCutchan, F.; Vodkin, M.; Lepia, S.H.; Schmidt, J.				
Gene 69, 287-300, 1988				
A:Title: Sequence and analysis of the DNA encoding protective antigen of Bacillus anthracis				
A:Reference number: I39933; MUID:89172073				
A:Accession: I39933				
A:Status: preliminary; translated from GB/EMBL/DBJ				
A:Molecule type: DNA				
A:Residues: 1-192 <RES>				
A:Cross-references: GB:M22589; NID:g143280; PIDN:AAA22636.1; PID:g143281				

```

Query Match      5.6%; Score 211.5; DB 2; Length 204;
Best Local Similarity 34.8%; Pred. No. 0.0002;
Matches 46; Conservative 31; Mismatches 34; Indels 21; Gaps 4;

Oy    616 MNILIRKRFHYDRNNINAVGADSEVYKEAREHREVINSTEGLLLNIDKIRKILSGYVEI 675
       ||||| : ||| | | : | : | : | : | : | : | : | : | : | : | : | : |
Db     1 MNILVRDP-YHYDNNGNIVGDVSYLEKNAYKOILMWSSDGVSLNDEPDNALSGYMLQI 59

Oy    676 EDTE-----GLKEVINDRYMDLNISSLRDQGFIFDEPKKNDKLPLYISPN 722
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     60 KKPSNMHLTNSPVTITLACKDSGCVGELEYRLS-----DGAGFLDENFKFDEMRSILV-DPG 112

Oy    723 YKVNYVAATKEN 734
       | ||||| :
Db    113 DDVVYYVAATKED 124

RESULT 4
G59104
hypothetical protein px01-111 - Bacillus anthracis virulence plasmid px01
C:Species: Bacillus anthracis
C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 11-May-2000
C:Accession: G59104
R:Okamoto, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehler
J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of px01, the large Bacillus anthracis plasmid harbored
A:Reference number: A59091; MUID:99445483
A:Accession: G59104
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-204 <OK!>
A:Cross-references: GB:AF065404; NID:g4894216; PIDN:AAD32415.1; PID:g4894327
A:Experimental source: strain Sterne
A>Note: similar to hypothetical protein in the protective antigen domain; ypa, plasmid F
C:Genetics:
A:Gene: px01-111
A:Genome: plasmid

Query Match      5.6%; Score 211.5; DB 2; Length 204;
Best Local Similarity 34.8%; Pred. No. 0.0002;
Matches 46; Conservative 31; Mismatches 34; Indels 21; Gaps 4;

Oy    616 MNILIRKRFHYDRNNINAVGADSEVYKEAREHREVINSTEGLLLNIDKIRKILSGYVEI 675
       ||||| : ||| | | : | : | : | : | : | : | : | : | : | : | : | : |
Db     1 MNILVRDP-YHYDNNGNIVGDVSYLEKNAYKOILMWSSDGVSLNDEPDNALSGYMLQI 59

Oy    676 EDTE-----GLKEVINDRYMDLNISSLRDQGFIFDEPKKNDKLPLYISPN 722
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     60 KKPSNMHLTNSPVTITLACKDSGCVGELEYRLS-----DGAGFLDENFKFDEMRSILV-DPG 112

Oy    723 YKVNYVAATKEN 734
       | ||||| :
Db    113 DDVVYYVAATKED 124

RESULT 5
F82865
hypothetical protein UUA42 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: F82865
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A>Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mild
A:Reference number: AB2870
A:Accession: F82865
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-468 <GLA>
A:Cross-references: GB:AF002145; GB:AF222894; NID:g6899476; PIDN:AAF30894.1; GSPDB:GN001
A:Experimental source: serovar 3, biovar 1
C:Genetics:
A:Gene: UUA42
```

[illegible]

A:Reference number: A86625
A:Accession: A86827
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1072 <STO>
A:Cross-references: GB:AB005176; NID:g12724625; PIDN:AAK05715.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yq16

Query Match 5.0%; Score 187; DB 2; Length 1072;
Best Local Similarity 18.8%; Pred. No. 0.043;
Matches 146; Conservative 151; Mismatches 333; Indels 148; Gaps 29;

```
QY 7 LIPALSTIIIVSSTGNEVIOAEVQKQENRLNESESSOGLLGYFSDNFAQPMVYTS 66
DB 319 LPISEASSVDMNTLNLSSLDSSIS-----SQTENSSG-----A 355
QY 67 STGDLSPSSLENIPEBENQFQSAISGFIKKVKSDEYTFATADNHVYMWVDQEVY 126
DB 356 SSTAIISYDSENSLSNSNO-----INSNSSEKDSNQSSLSGSSMSSESEHSNSNI 410
QY 127 NKASNSNKR-----LEKGRLYQIKIYQORENPEKGLDFKL-----YWTDSQNKREVISS 177
DB 411 NETNNSSEITNLPRSPNPTESNSVSDQTSSEASTNSNSISLSPSNISSTSDSEATNS 470
QY 178 DNLQLELKQKSSNRKRKSTSGPTVPDRDNDGI-----PDSLEVEGYTVDKNRRTFLS 233
DB 471 DFNVAEVAENNSLAVNSSSSVLSSTADNLGINOGSDNLTKRD-----SPEISTGAFLS 528
QY 234 PMISNIHKKGLTKYKS-----SPEKWS-----TASDPYSDFKYNVGRIDKNVSPKAR 281
DB 529 ---SNQTSSEASTNSNSISLSPSNISSTVLESTTSSNFSNVAEVAENNSLAVNSSS 585
QY 282 HPLVAAYPIVHVD-----ENILSKNEDQ-----STQNTSETRTISKNTSRTHTS 330
DB 586 SVLSSTSTADNLEINQFSDNITKDSSELTSGAFLSNQTSSEASNSMSINSPLSL 645
QY 331 EYHGNAEVAHANTSTSRHTSEYHGNAEVAVAIDHSLSLAGERTVAETWGLTADTARLN 390
DB 646 SLTNSSEATNOSNS-SEATKYDNNSSTHSSNLTNS-----GSDSDSDSDSDSDSSNL- 699
QY 391 ANIRVNTGTAPIYVNLPTSLVLCNOTLTIKAKENQLSQ-----ILAPNNYYSKMLA 446
DB 700 -----SSPNLETTNOTIISKPEVNNISNPCKVSSNSNVQENSTDH 741
QY 447 PIALNAQDFFSPTITMYNOFLEKTKQLRLDQVYGINATYN-----FENGRRVAVDTG 503
DB 742 EMSTPKRSISSPISTSSQKESQSN---LLNTTEGINNPITFNSSSEASASILT- 797
QY 504 SNMSEVLPQIOETTARIIF-----NGKDL-----NLVERRIAVNDSPLETTKPD 550
DB 798 -SYSNNSESSSTGCLYISNEAQRNGSEISHLSPSSNSNENNVASSISOALLESKST 856
QY 551 TLK-FALIAFGFNEP-NGNLQYQKDTPEFPNFDQOSONIKQALAEANTNYTYVD 608
DB 857 NKRSSSLIINSTHPQEDNOSNSD-----EVSNNVSEIILQGLNISKTHN 908
QY 609 KI---KLAKNNILIRDRFRHYDRNNIIVGADSEVYKAEHREINSSTEGLLANDDKDIR 665
DB 909 SLTSGKLSVITLPSKSVYTNKNSNMVSEKLIKTPQKN-DESONLGOJLTALDSFN 967
QY 666 KILSGLYIEIDTEGLKEVIN-----DRYDMLNSSLROD-GKTPIDEKKYVDKL 714
DB 968 K-----EVEETMEDSKTVPDKVLNDENGDRSQNNKTSTIAKDKKVFKEKRFSEFSKI 1018
```

RESULT 7
553805
alpha-toxin - Clostridium novyi (ATCC 19402)
C:Species: Clostridium novyi
A:Variety: ATCC 19402

C:Date: 28-Oct-1996 #sequence_revision 08-Nov-1996 #text_change 05-Nov-1999
C:Accession: S55805; S71294; S71158; S44273; I40834; S44272
C:Hoimann, F.; Herrmann, A.; Habermann, E.; von Eichel-Streiber, C.
Mol. Gen. Genet. 247, 670-679, 1995
A:Title: Sequencing and analysis of the gene encoding the alpha-toxin of Clostridium
A:Reference number: I40834; MUID:95342160
A:Accession: S55805
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-2178 <HOF>
A:Cross-references: EMBL:Z48636; NID:g728537; PIDN:CAA8565.1; PID:g755724
A:Accession: S71294
A:Molecule type: protein
A:Residues: 1-15 <HOW>
A:Hoimann, F.
Submitted to the EMBL Data Library, March 1995
A:Reference number: S71158
A:Accession: S71158
A:Molecule type: DNA
A:Residues: 1-1179, 'LKV', 1183, 'LVTHIGE', 1191-2178 <HOS>
A:Cross-references: EMBL:Z48636; NID:g728537; PIDN:CAA8565.1; PID:g755724
A:Hoimann, F.; Habermann, E.; von Eichel-Streiber, C.
Submitted to the EMBL Data Library, July 1993
A:Description: Sequence analysis of Clostridium novyi alpha-toxin: a member of the fa
A:Reference number: S44272
A:Accession: S44273
A:Molecule type: DNA
A:Residues: 1-243; 1204-2178 <HON>
A:Cross-references: EMBL:Z23281
C:Genetics:
A:Gene: tcn-alpha
C:Superfamily: cpl repeat homology
C:Keywords: virulence factor
E:1880-1899/Domain: cpl repeat homology <CP12>

Query Match 4.9%; Score 183.5; DB 2; Length 2178;
Best Local Similarity 20.6%; Pred. No. 0.18; Mismatches 241; Indels 235; Gaps 34;

```
QY 122 DOEVINKASNSNKRILKGRLYQI-KIYQOREN-----PTEKGLDFKL----- 163
DB 202 DDIINFLSNFK-----YDIKLNKNKNNKKAIGATQINENILTNKLSGY 253
QY 164 YWTDSQNKREVISSDNLQLELKQKSSNRKRKSTSGPTVPDRDNDGIPDSLE-VEGYT 222
DB 254 YOELIOTNNLAAASDILRIALILKKG-----GVYCDLDFLPQVN 292
QY 223 VDVKRKRTFLSPWISN-----IHEKKGLTKYKSSPEKWSFASDPYSDFEKVGR 272
DB 293 LSLFNDISKPNQMSNYEALFEALIANDEKLMNY-----PKYMEQVPSSEI 340
QY 273 DKNVSPKAR---HPLVAAYPIVHVDENILSKNEDOSTQNTDSETRTISKNTS-TSRT 327
DB 341 KERILSFVRNHINDLILPLGDIKISOLEILSLRIKATGKKTFSNAPLISNDSLFTLN 400
QY 328 HTSEVHG-----NAVYHANTSTSRHTSEYHGNAEVAVAIDHSLSLAGERTVAETMG-L 381
DB 401 LISOLENREYELINSIQEKFKICEFYDSYINSSELVETTPKNSLMDGSSFYQOIIGYL 460
QY 382 NTADTARNANIRY-----VNTGTAPIYVNLPTT-SLVGKNQTLATIKAKENQLSQILA 435
DB 461 SSGFPEVNSVYFESGPNITSSATCDTYHFKINTDMLSSQOEI-----FEA 508
QY 436 PNNYPSKRLAPIALNAQDFFSPTI---TNMYNOFLEKTKQLRLDQVYGINATYN 492
DB 509 SNNLFSK-----THDEFKSSWLSNIAEKFOQLIKITYIGR-----TLN 549
QY 493 FENGR-----VRVDTSNMSEVLPQIOETTARIIFNGKDLNVERRIAVNDSPLETTK 547
DB 550 YEDGINFKKMKRVTP-----SELLKYLIEVNSPKIYENDLNNI-----LQIQG 593
QY 548 PDMTLKALKIAFGFNERNGNLQYQKDTPEFPNDQ---QTSQNIKQALAEIN----- 599
```

[illegible]

RESULT 8
S31262
Tyb protein - yeast (*Saccharomyces cerevisiae*) retrotransposon Ty4 (fragment)
C:Species: *Saccharomyces cerevisiae*
C:Date: 18-Jun-1993 #sequence_revision 25-Apr-1997 #text_change 26-May-2000
C:Accession: S31262
R:Janetzky, B.; Lehle, L.
J. Biol. Chem. 267, 19798-19805, 1992
A:Title: Ty4, a new retrotransposon from *Saccharomyces cerevisiae*, flanked by tau-elements
A:Reference number: S31261; MUID:93015829
A:Accession: S31262
A:Molecule type: DNA
A:Residues: 1-1465 <JAN>
A:Cross-references: EMBL:S46865
C:Genetics:
A:Mobile element: retrotransposon Ty4
;Superfamily: Tyb protein

```

Query Match      4.9%; Score 183; DB 2; Length 1465;
Best Local Similarity 20.1%; Pred. No. 0.11;
Matches 167; Conservative 117; Mismatches 283; Indels 264; Gaps 40.

OY      25 EVIGAEVQKQENRLNLESSESSOGGLGYFSDLNFQAPMYWTSSTTGDISIPSELENIPS 84
       +::+::+: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      735 ELIQEQLKKTN---HETSPFKEGSIG---TNKFQNTNNELSKTGDSTLP+IKTLTESINN 788

OY      85 ENQYFOASIMSGFIKKVKSDEYTFATSDNHVTMNVDDQEVINKASSNKIRLEGRGLYO 144
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      789 HH-----SNDY-----STNKVE-----800

OY      145 IKIYOOREN----PTEKGIDFLKYTWDSQONKKEYISSDNLQELPKOKSSNRKRRTSA 200
       ::::: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      801 ---KFEKNHNHPRIEDIVDM-----SQOTMESMCQODNNKEKLKVYDKNPPTONGTV 852

OY      201 GPVPVRDRDNDSIPDSLEVEGYTVDVKNKRTELSPWISNIHEKK-----GLTKY--- 248
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      853 SPR-----LEQNIIEAGSPVOYVNKSAFLNKFESSLNMKRRKHHDKNNSLTSELE 904

OY      249 -----KSPFEKKSTASDPYSDFEKTGTGRIDKNVSPARHPPLVAAYPTLVHVMEN 297
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      905 RDKRRSKKNRVKLLIPDNNETVSAPKIRAIIYNEAISKNDPLEKEHYYKOAY--HKELQN 961

OY      298 IILSKNEQOSTONDSETRTISKN--TSPSRHTSEVG--NAEYHANSTSRTHSEVH 353
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      962 LKDKMKVEDVDYKGSSE---IPDNLIIVPTNTITFKRKNGITYARIYACRGDTOSPDYYSVI 1018

OY      354 GNAEYHAVALIDSLSLAGERT-WAETGCLINTADT-ARLANANTRYVNTGAPLYYNVLPR-TT 410
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1019 TTESLNHHHIKIPLMIANRNRMFMKTLDIINAHFLAKLEE-----IYIPPHDR 1068

OY      411 SLVLRKNOTLATIRKKEQO-----LSQT-LADPNNTYP-----SKULAPIALNAQDD 455
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 1069 RCVAKLMAKLYGKQSPKEMNOHLRQYLANGIGKONSYPGLYQEDBKML-IAVYDCC 1127
QY 456 FSSPTITMNTNOFL-ELEKTKOLR-----LTDQVYGNIAFYNEBNGV----- 498
Db 1128 VIASANEORLDREFINKLSNFEKLTIGTLIDDLVDLD-ILGMDLYNNRLGTIDLTKEF 1186
QY 499 --RDTGSNNMSEVLPQLOETARILFNGCDLNIYERIAAMP-SDPLETTKPD-----M 550
Db 1187 INRND--KKYNEELKTRKSSIPHMSTYK-----IDPKKDVLQMSSEEPQCVL 1233
QY 551 TLKEALKIAGFNEPNNGNIDYOGKDLTEPEFNFDQOTSQIKNOIAELNATNYYVLDKI 610
Db 1234 KLOOLL-----GELNYV-RHKCRDIDIEPAKKAVKARLVNHYHERFYWTIKYLIOTL 1282
QY 611 KLNKKNMILIRKRFHYDN-----NIAGADESVYKEAHR-----VI----- 649
Db 1283 -----VRKKDIGIHDRCDNCKDKVIAITDASVGEYDAQSRIGVILWGMNIFNYS 1335
QY 650 NSTTEGLLINDKDIRKILSGY-----IYELIEDT-----BSLK--- 682
Db 1336 NKSJNRCSVSTEAEHLAIFYGADSETLVYTLKELEGDNNDIWMITSPKALQIGLNSRY 1395
QY 683 -----EYINDRYMNLISSLRODGTFPI-----DPKKY 710
Db 1396 QQPEKFTWIKETELIKELIKESIKILKLTGTGNGNIDLLTKRVASDPDRF 1446

```

RESULT 9
S56894
TYB protein - yeast (*Saccharomyces cerevisiae*) retrotransposon Ty4.JL
N:Alternate names: protein J0780; protein YJL13w
C:Species: *Saccharomyces cerevisiae*
C:Date: 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 17-Mar-2000
C:Accession: S56894
R:Czlepuch, C.; Kordes, E.; Pujol, A.; Jauniaux, J.C.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56891
A:Accession: S56894
A:Molecule type: DNA
A:Residues: 1-1803 <CZ1>
A:Cross-references: EMBL:Z49389
C:Genetics:
A:Map position: 10L
A:Module element: retrotransposon Ty4.JL
C:Superfamily: TyB protein

```

Query Match Similarity      4.9%; Score 183; DB 2; Length 1803;
Pred Local Similarity      20.1%; Pred. No. 0.15%;
Matches 167; Conservative 117; Mismatches 283; Indels 264; Gaps 40.

QY      25 EVIAQVAKOENRLNLESESSOGGLGYYFSDLNFOAPRWVYTSSTGDISIPSELENIPS 84
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1073 ELIQOLKKTN---HETGFPPKGSIG---TNYKFRNTNNEISLKTGDTSLPIKTELSINN 1126

QY      85 EMQYQSAIWSGFIKVKKSDEYTFPTASDNHTMNVDDQEVINKSNSKIKILEKGRLYO 144
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1127 HH-----SNDY-----STNKV-----1138

QY      145 IKIQYQREN---PYEKGIDFLYMTDSQNKKEVISDNLQJPELKOKSSNRKKRSTSA 200
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1139 ---KKEKENHRRPREDLYDM-----SDQTOMESNOQGGNNLKLKVDKNPPTDNGTNV 1190

QY      201 GPTVDRDRNDGIPRDSLEVEGTYVDVKKRRTGLSPMYSNIHEK-----GLTKY--- 248
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1191 SPR-----LEQVIEKSGSPVQTVNNSAFNLKEFSSILNMKRKKRRHDKNNLSLYELE 1242

QY      249 -----KSPSEKASTADPYSDPEKYVGRIDKNVSPARRPLVAAYPIVADNMEN 297
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1243 RDKKRSKKNRKYLIDNNETVSAPIRATYYEALISKNNDLKEKHIEYKQAY---HKELON 1299

QY      298 IILSKNEDOSTONTDSEFTTITSK---TSTSRTHTSEVHG--NAEYHANTSTISRTHTSEVH 353
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```
Db 1300 LKDMKVPDVVYSRSE---IPDNILVPPNTITFKRNGITAKIAYCQGDOSPDTYSVI 1356
Qy 354 GNAEVHAYVAIDHSLSLAGERT-WAETMGLTADT-ARLNANIRYVNTGAPLYNLP-TT 410
Db 1357 TTESIHNHHIKFIEMIANNRNFMFKTLIDINHAFLYAKLEEE-----IYIPHPDR 1406
Qy 411 SLVLCNKTATIKKKNQ-----LSQI-LAPNNYF-----SKNLAPIALMAQDD 455
Db 1407 RCVVNLNALYGLKQSPKEMNDHLROYLNGIGLKDNSTYPLGYOTEDKILM-IAVYVDDC 1465
Qy 456 FSSPTITNNYNOFL-ELEKTOLR-----LPTDQVYGNIAIYNGENGV----- 498
Db 1466 VTAASNEQRLDEFINKLSNFKLITGTLIDVLDFTD-ILGMDLYYNNKRLGIDTLKSF 1524
Qy 499 --RVDTSNMSEVLPQIOETARIITFNGKDLNVERRIAANVP-SDPLETTKPD-----M 550
Db 1525 INRMD--KKYNEELKIKRSISIPHMSYTK-----IDPKDVOLOMSEEPERQVYL 1571
Qy 551 TLKELKLAIFGNEPENGULQYOGKDIPEDFNPDQOTSNIKQLAELNATNIYVLDKI 610
Db 1572 KIQQLL-----GELNTY-RHKCRYDIEFAVKVAKARLVNYPHERVYMIYKIQYL 1620
Qy 611 KLNAMNILLIRKREHYRN-----NIAVGADESVYKREHRE---VI----- 649
Db 1621 -----VRKDIGIHYDRDCNKKKVAITDASVGEYDAQSRIGVILMYGMNIFNYS 1673
Qy 650 NSSTEGLLNIDKIRKLISGY-----IYIEEDT---EGIK--- 682
Db 1674 NKSTNRCSVSTAEELHAIYEGADSEFLKVLKELGEGDNNDIVMTJDSKPAIQGLNSY 1733
Qy 683 -----EVINDRYDMNITSLRQDGKTF-----DRKKY 710
Db 1734 QOPKEFTWIKETIIEKIKESIKLKITGKGNITADLTKRVASASDEKRF 1784

RESULT 10
T18351
Impl protein - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 07-Dec-1999
C:Accession: T18351
R:Jensen, L.T.; Ladefoged, S.; Birkelund, S.; Christiansen, G.
Infected. Immun. 63, 3336-3347, 1995
A:Title: Selection of mycoplasma hominis Pg21 deletion mutants by cultivation in the pre
A:Reference number: Z18884; MUID:95369882
A:Accession: T18351
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Cross-references: EMBL:U21961; NID:g790241; PID:g790242; PIDN:AAA81012.1
C:Genetics:
A:Gene: Impl1
A:Genetic code: SGC3
```

```
Query Match 4.8%; Score 179.5; DB 2; Length 1051;
Best Local Similarity 19.1%; Pred. No. 0.1;
Matches 152; Conservative 123; Mismatches 324; Indels 195; Gaps 30;
```

```
Qy 19 SSTGNLEVIOAEVQENLNLSESSSOGILGYFSDLNFOAPMVYTS---TTGDSLTP 75
Db 320 SNKSDIESANTELKQALAKANADKVOADNLAKSIKEQLNNSVSNANTLSAKLTDXDNTIQ 379
Qy 76 SSELNIPSENOYFOSAIWSGFIKKKSDERTFATSADNHVTMMVDDQEVINKASNS--N 133
Db 380 QAKTE-LEKEIQKANQATKSNNTTASMSQAS-----SSLDAXVALEITKKKLETFKDKDEAFN 434
Qy 134 KIRLEKGLYIKIOYQRENPTKGL-----DFKLYWDSQKKKEVISSDNILQPEL 185
Db 435 ELKQTRNOIOEF-INTNNKNPNYSSELLISQLTNSKRDSKNSVTDSSNKSIDIESANT---EL 489
Qy 186 KOKSNSNKKRSTASGPIVPPDRNDGIPDSLEVEGYTVDVKNKRFLSPWISNHE----- 241
```

```
Db 490 KOALNTAKAKSSI-----DNELRPLKNDLQSKIEEFGIRINTNFSWISSKLETTKN 541
Qy 242 --KKGLTK--YKSSPEKWSFA-SDPYDSEKVTGRIDKNVSP-----ARH 282
Db 542 KLAELTLDADAKNNPSSSKQALKKSSQOYOKLGNELKLTIEEGKVEYETKNSNGYRLF 601
Qy 283 PLVAAPVPIVHDMENI-----ILSKNEDQSTQNTDSETRITSKNTSTSRHTSEVH 333
Db 602 KLAQAEQFNNSDVKLKNAMWEKQTLISKQKLGQNSTKDYLTQLSTEMSTQESTIKKVI 661
Qy 334 GNAEVHANTSTSRHTSEVHGAEVHAYVAIDHSLSLAGERTMAEWMG---INTADTAR- 388
Db 662 VNIQAHIRNNLSQYRLERADKLIAMKKGQYGDVIESLQKWQMDLSDSVSDSLKDD 721
Qy 389 LMANIR-----YVNTGAPYVNLPTTSLVLCNKTATIKAKE----- 427
Db 722 FNKARVLVGDYTKNPYSSWFINKRNSIEYQULRLILVREHEIILLDKAKDDKRAE 781
Qy 428 -----NOLSOILAPNNYPSKNLAPIALNAQDDFSSPTITMANYOFLERKTOL 477
Db 782 KTIKFDENINSIDQ-----RAKRILQOELINKAKNDLSNFTLNHQKNOFTAKDITPKI 833
Qy 478 RUDTQVYGNIAIYFENGVRVVDGSMSEVLPQIOETT---ARIIFNGKDLNVERR 533
Db 834 SL-----LEKKLNEIN---OYLLPIIKRAVSKISEIKRKKLEEDIR 875
Qy 534 IAAVNSDPLETTKPDMLKEALKIAFGNEPENGULQYOGKDIPEDFNPDQOTSNIK 593
Db 876 NFYLMEKEVINKYISELNNKY-----ELRSININEK-----WESTKD 914
Qy 594 QLAELNATNIYVLDKIKLNAMK--NILLRDKREHYDRNNIAVGADESVYKREAH----- 645
Db 915 TLDNIN-----IKENVSLLKEVIITNNSNAQYSINRIISTVPE-FIKVAQTTRSNN 963
Qy 646 -REVINSSTEGLLNIDKIRKLISGYIYIEEDPGLKEVINDRYDMNITSLRQDGKTF 704
Db 964 LRSIAEQFKPLTLDIDKNLKEV-----KKIIDEN-----KTLNSD----- 999
Qy 705 IDFKRYNDKRLPLYI 718
Db 1000 -NIAKLNEKIRLVL 1012
```

```
RESULT 11
T30822
Impl protein - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 07-Dec-1999
C:Accession: T30822
R:Jensen, L.T.; Ladefoged, S.; Birkelund, S.; Christiansen, G.
Infected. Immun. 63, 3336-3347, 1995
A:Title: Selection of Mycoplasma hominis Pg21 deletion mutants by cultivation in the
A:Reference number: Z18884; MUID:95369882
A:Accession: T30822
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1365 <JEN>
A:Cross-references: EMBL:U21962; NID:g790243; PID:g790244; PIDN:AAA81013.1
C:Genetics:
A:Gene: Impl1
A:Genetic code: SGC3
```

```
Query Match 4.8%; Score 179.5; DB 2; Length 1365;
Best Local Similarity 19.1%; Pred. No. 0.15;
Matches 152; Conservative 123; Mismatches 324; Indels 195; Gaps 30;
```

```
Qy 19 SSTGNLEVIOAEVQENLNLSESSSOGILGYFSDLNFOAPMVYTS---TTGDSLTP 75
Db 634 SNKSDIESANTELKQALAKANADKVOADNLAKSIKEQLNNSVSNANTLSAKLTDXDNTIQ 693
Qy 76 SSELNIPSENOYFOSAIWSGFIKKKSDERTFATSADNHVTMMVDDQEVINKASNS--N 133
```

Db 694 OAKTE-LEKEIOKANOAIKSNMTASMOQAK-----SSLDAAKVAEITKLETFNKKKEAKFN 748
QY 134 KIRBLEGRILYOIKIOYQRNPTPEKL-----DPKLTWTSQNNKEVYSSNLOLDEL 185
Db 749 ELKQTRNQIOER-INTNKNPNYSELISQLTJSKRDSKNSVTSNNDIESANT-----EL 803
QY 186 KQSSNSRRKRSTSAQPTVPDRDNDGIPDLSLEVEGYTVDKNNKRTFLSPWISNIHE---- 241
Db 804 KQALNTAKAKKSSI-----DNEIRPLKNDLOSKIIEFGPIRANPNFMSISKLETTKN 855
QY 242 --KKGITK--YKSSPEKMTA--SDPYDFEYVTRGRIDKNVSE-----ARH 282
Db 856 KLAELITLKADAIKNPNSSKQALKDSSQOVOKLGNELLKTIPEFGKVEETKNSNIGYRLF 915
QY 283 PLVAAYPIYHVDMEN-----ILSKNEDOSTONTSERTTICKNTSTSTHTSEVA 333
Db 916 KLAQABOFNNSQVQDKKNAMEEKQTLISKQKLGNOSTDYLTQLTSTEWSTOSTIKKVI 975
QY 334 GNAEYHANTSTSRHTSEVHGNAEVAHAVIDHSLSLAGERTMAETMG---LNTADTAR-- 388
Db 976 VNIQAHIRNNLSQYRLKADKLIANKRGYGVKGLIESIQKQMDLMDSDVLSVDDSLKD 1035
QY 389 LMANIR-----YVNTGTADPIYVNLPTTSLVLGKNQTLATIKAKE----- 427
Db 1036 FPKALRVLVGDYTKNPNVSSWFIKNRNSIENYQNLRLILVRENEILLDKAKDLDRAE 1095
QY 428 -----NOLSOILANNYPPSKNLAPIALNODDSSPTITANNYQFLEETKOL 477
Db 1096 KTIKFEVDENINSLDQ-----RAKRLKQELIANKNDLSNFTLNHQKQFATAKDTTPKI 1147
QY 478 RLDTDOVYGNIAITYNENGRVVRDTGSNMSEVLPJOIETT---ARLIFNGKDLNIVER 533
Db 1148 SL-----LEKNLINEIN-----QYLLPITKEKAVAKISIELEKKNKLEEDITIS 1189
QY 534 IAAVNPSPDLETTKPDMLTKLEALKIATFGNEPNGLQYOGKDTPEDFNFDOOTSONIKN 593
Db 1190 NFYLMEKVEINIKYISELTNKKY-----ELRSNINFEKK-----WESIKD 1228
QY 594 QLAELNATNIYVLDKIKILANAKM--NILRDKRFHIDRNNIAVGADESYYKEAH----- 645
Db 1229 TLDNNI-----IKENVSILKEVIINNSNAOYSIKRILSTVEP-FIKAQOTTRSN 1277
QY 646 -REVINSSTEGLLNIDKDIRKILSGYIVIEIDTEGLKEIYNDRYOMLNISSRODGKTP 704
Db 1278 LRSLEIOLKQFPLDIDDKMLKEY-----KKILDEN-----KTLNSD---- 1313
QY 705 IDPKRYNDKRLPLYI 718
Db 1314 -NIAKLNEKIRLLV 1326

RESULT 12
PC1253
TYB protein - yeast (*Saccharomyces cerevisiae*) retrotransposon Ty4 (fragment)
N:Contains: integrase; proteinase; reverse transcriptase; RNase H
C:Species: *Saccharomyces cerevisiae*
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-May-2000
C:Accession: PC1253; S27465
R:Stucka, R.; Schwarzlose, C.; Lochmuller, H.; Haecker, U.; Feldmann, H.
Gene 122, 119-128, 1992
A:Title: Molecular analysis of the yeast Ty4 element: Homology with Ty1, copia, and plan
A:Reference number: J01482; MUID:93083972
A:Accession: PC1253
A:Molecule type: DNA
A:Residues: 1-1467 <STU>
R:Stucka, R.; Schwarzlose, C.; Lochmuller, H.; Hcker, U.; Feldmann, H.
submitted to the EMBL Data Library, May 1992
A:Description: Molecular analysis of the yeast Ty4 element: homology with Ty1, copia, an
A:Reference number: S27465
A:Molecule type: DNA
A:Residues: 1-1467 <STW>
A:Cross-references: EMBL:M94164

C:Genetics:
A:Gene: TY4B
A:Mobile element: retrotransposon Ty4
C:Superfamily: TyB protein

Query Match 4.7%; Score 177.5; DB 2; Length 1467;
Best Local Similarity 19.7%; Pred. No. 0.21;
Matches 154; Conservative 121; Mismatches 277; Indels 231; Gaps 38;

QY 25 EVIOAEVQKQENRLNSESOGILGYFSDLNFOAPMVYTSSTGDLSPSSLENIPS 84
Db 736 ELIQEDQKKTN---HETSPFEKESIG---TNVPRNNTNNEISLKTGDFSLPIPTLESINN 789
QY 85 ENQYFOSAIWSGFIIKAKKSEDTFATSADNHVTMVDQEVINKASNKIRLEKGLYQ 144
Db 790 HH-----SNDY-----STNKEV----- 801
QY 145 IKIYOQREN-----PTEKGLDFKLYWTDQNNKEVYSSNLOLDELAKQSSNSRRKRSTSA 200
Db 802 ---KFEKENHHPRIEDIVM-----SDQIDMESNCQDGNLKEIKLYTDKNNVPDNGTNY 853
QY 201 GPTVPDRDNDGIPDLSLEVEGYTVDKNNKRTFLSPWISNIHEKK-----GLTKY-- 248
Db 854 SPR-----LEQNIERSGSPVQTVNKSAPLNKEFSLNKKRRKRHDKNNSLTSEYE 905
QY 249 -----KSSPEKMTASDPYSDFEVYTRGRIDKNVSEPEARHPLVAAYPIYHVDMEN 297
Db 906 RDKRSKKNRVKILPDIMETVSAPKIRAIYVNEAISKNPDLKEKHEYKOAY---HKELON 962
QY 298 IILSKNEDOSTONDSE---TRTISKNTSRTHTSVHGNAEYHANTSTSRHT----- 349
Db 963 LKDMKVFEDVDKYSRSFIPDNLIVPTNTIFTKKNRGITKAKRIYCRGDTGSPDYTSVITTE 1022
QY 350 SEVHGNAEVAHAVIDHSLSLAGERTMAETMGTLADT-ARLNNANIRVYNTGTAPIYVNL 408
Db 1023 SLNHNHNIKILMQOTTEICL-----W--TLDINHAFIYALKEELTYIPRHLIGGVYKYL- 1074
QY 409 TTSVLVGNKQTLATIKAKENQ-----LSQI-LADNNYTP-----SKNLAPIALNAQ 453
Db 1075 -----NKALYGLKQSPKEMNDHLROYLNGIGLKQNSYTPGLYQTEDNKLAM-IAVYVD 1125
QY 454 DDESSPTITANNYQFL-ELKTKQLR-----LDTDOVYGNIAITYNENGRV----- 498
Db 1126 DCVIAASNEQRLDEFFKILSNFELKITGTLIDVDLTD--ILGMDLYVNRRLGTIDITLTK 1184
QY 499 ---RVDTGSNMSEVLPJOIETTARLIFNGKDLNLYERRIAAVNP--SDPLETTKPD--- 549
Db 1185 SFINRMD--KKYNDELKIKKSSIPHMSTYK-----IDPKKDVLOMSEEPEROG 1231
QY 550 -MTLKEALKIATFGNEPNGLQYOGKDTPEDFRFDQOTSONIKQLAELNATNIYVLD 608
Db 1232 VLKLOQL-----GELNYY-RHKCRYDIEFAKKAARLVNYPHEVFYVIYKIIQ 1280
QY 609 KIKLANKMNILIDDKRFHYRN-----NIAVGADESYYKEAHNR-----VI----- 649
Db 1281 YL-----VRYKDIGHYRDCKKDKKYAITIDASVGEYDAQSRGIVLWGMNIFNV 1333
QY 650 --NSSTEGLLNIDKDIRKILSGYIVIEIDTEGLKEIY-----NDRYDMNIS---SLRQ 699
Db 1334 YSKNSTRCVSSYAEALHAIYEGV---RDSFTLKVTLKELGEGDNDNDYIMVITVKVYLKE 1389
QY 700 DKG 702
Db 1390 LGE 1392

RESULT 13
SS5101
hypothetical protein YMR219w - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: hypothetical protein YMR261.13; hypothetical protein YMR959.01
C:Species: *Saccharomyces cerevisiae*
C>Date: 08-Jul-1995 #sequence_revision 09-Mar-1996 #text_change 29-Oct-1999

Db 612 LGAISLINKKENSMPITESYEIPNDMLGLPLNDLNEKLEFNISYKNTAYEFKKIYFNEDQ 671
Qy 410 -----TSLVLGNQTLA---TKAKENQSLIAPNNYPSKRLPIALNA----- 452
Db 672 WMTQYQYQFDLIGMARSRVLAQETLIRIKLOKLSYLIGSNISSDNLALMLTTNTL 731
Qy 453 QDDSSPTITMNN-----YNQFLELEKTKQLRLDDQVYG--NIATYN 492
Db 732 RDISNEQIAMNNVDSFLNNAALCVFESNIPKIFSP-----MEQINNINIKTKE 782
Qy 493 FENGRVAVDTGSNMSEVLPQIETTARIIFNGKDLNVERRIAANVSDPLETTKPPMTL 552
Db 783 F-----IOKCTININDEKQLQNLN-----VENSIDFEELN-----IQMKSLSESETALLI 829
Qy 553 KEAL-----KIAFGNENGN-----LOYQKDI-TEPPND-----QQTSON 590
Db 830 KEETWPELVLYAFKEGNNVIGDASGKNTSIEF-SKDIGLYVINDALYLNOSIS 888
Qy 591 IKNQLAELNATNIYTV-----LDKIKLNAM----- 616
Db 889 FSNDFENGSLNFSSTIFWMLNLOKDTIKSLKSGKEDNCWEIYFODTGLVFNKIDSNG 948
Qy 617 ---NILIRD---KRFHY-----DR--NNIAGADESVY-KEAHREVINSSTEGLLNIDK 662
Db 949 NEKNIVSDVSNNSMWHITTSVDRLEKQOLLIFIDDLVANSIKELINITSNIIISLSE 1008
Qy 663 DIRKILSGYVEIEDTGLKEVINDRYDMLNIISSLRDQKTFIDPK-----YNDKL 714
Db 1009 NNSYIGLITLKNPPTS-QEVLISNFEVLNNSYIRDSNEERLEYNKTYQLYNYVFSDPK 1067
Qy 715 PLYISNPNKYVNYAVATKENT 735
Db 1068 ICEYKQNN---NIT-LTINNIT 1084

RESULT 15
S52611
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty4
N/Alternate names: protein YHL008v-a
C/Species: Saccharomyces cerevisiae
C/Date: 05-May-1995 #sequence_revision 19-Oct-1995 #text_change 23-Mar-2001
C/Accession: S52611
R/Favella, T.
submitted to the EMBL Data Library, June 1994
A/Description: The sequence of S. cerevisiae cosmid L5018.
A/Reference number: S46798
A/Accession: S52611
A/Molecule type: DNA
A/Residues: 1-1802 <FAV>
A/Cross-references: EMBL:U01581
C/Genetics:
A/Map position: 8L
A/Mobile element: retrotransposon Ty4
C/Superfamily: Tyb protein

Query Match 4.68; Score 175; DB 2; Length 1802;
Best Local Similarity 20.1%; Pred. No. 0.38;
Matches 168; Conservative 118; Mismatches 276; Indels 274; Gaps 42;

Qy 25 EVIOAEVYKQENRLNSESOGILGYFSDLNFOAPVAVVSTTGDLSTPSSELENPS 84
Db 1072 ELIIO---EQFKTNHETSPKESIG---TNVFRNTDNETSLKTGDTSLPIKTLSEINN 1125
Qy 85 ENQYFQSAIWSGFIKVKSDYEFTATSADNHVTMMVDDQEVINKASNSKIRLEKGRLYQ 144
Db 1126 HH-----SNDY-----STNKVE----- 1137
Qy 145 IKIYOQEN---PTEKGLDFKLYWDSQNKKEVYISSDNLQLPKLOKSSNSRKKRSTSA 200
Db 1138 ---KFEKENHHPPIEDIVDM---SDQDTMESNCQDGNMKELKLYDKNVPTDNGTNV 1189
Qy 201 GPVTPDRNDQGIPOSLVEGYTVDVVKNKRFPLSPWISIHKK-----GLTKY--- 248

Db 1190 SPR-----LEQNIKESGSPVQTVNKSAPLNKFESSLNKKRRKRRDKNNSLTSVELE 1241
Qy 249 -----KSSPEKSTASDPYSDEKYTG-----RIDKNVSPKARHPLVAAYPIVH 292
Db 1242 RDKRKRKRNRKRLIPDNNETVS-----AQKIRAIYNEAISKNPDLKEKHEKQAY---H 1293
Qy 293 VDMENILSKNEDQSTQNTDSETRISKN---TSTSRFTSEVHG--NMEVANSTSRTH 348
Db 1294 KELQNLKMKYFDVVKTSRSE---IPDNLVPTWITFKRNGIYKARIYCRGDQSPD 1350
Qy 349 TSEVHGAENVAVAIDHSLSLAGERT-WAETMGLNTADT-ARLANIRYVVTGAPITNV 406
Db 1351 TVSVITTESLNNHNIKIFLMIANNRNMFMKLDINHAFVAKLEEE-----IYIP 1400
Qy 407 LP-TTSLVLGNQTLATIKAKENQ-----LSOI-LAPNNYV-----SKNLAPIAL 450
Db 1401 HPHDRRCVYKLNKALYGLKQSPKEWMDHLYOINIGIKDNSYTPGLYQTEDKNLM-IAY 1459
Qy 451 NAQDDFSTPTTMNNOFL-ELEKTKOLR-----LDTDOVYGNATYNEFNGRY-- 498
Db 1460 YVDCVIAASNEQRLDERFINKLSNFEKLTITGLIDVLDTD-ILGMDLVYNNKRLGTIDL 1518
Qy 499 -----RVDTGSNMSEVLPQIETTARIIFNGKDLNVERRIAANV-SDPLETTKPD- 549
Db 1519 TLKSPINRMD--KYNEELKIRKSSIPHMSTYK-----IDPKKDYLOMSEEF 1565
Qy 550 ---MTLEALKIAFGNEPNGNLQYQKDIITEFDNFEDQOTSCNIKQLAELNATNIYT 605
Db 1566 RGVLKLIQQL-----GELNYV--RHKCRYDINFAVKVAVRLVNYPHERVYMIYK 1614
Qy 606 VLDRKILAKANNILIRDKRFHYDRN-----NIAGADESVYKEAHRE---YI----- 649
Db 1615 ITQYVL-----VRYKDIGHYDRDCNKKRYALITDASVSEYDAQSRIGVILWYGMNI 1667
Qy 650 -----NSSTEGLLNIDKIRKILSGY-----IYEIEDT-----EG 680
Db 1668 FNVYSNKTNCVSTSEALAIYGYADSETIKATLKELEGENDIMTDSKPAIOG 1727
Qy 681 LK-----EVINDRYDMLNIISSLRDQKTFI-----DFKKY 710
Db 1728 LNRSYQPKREKFTWIKETIIEKIKIKESIKILKITGKGNIAIDLTKPYASADPKRF 1783

Search completed: December 2, 2001, 13:51:36
Job time: 331 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:49:34 ; Search time 55.64 Seconds
(without alignments)
297.267 Million cell updates/sec

Title: US-09-747-521-4_COPY_1_735
Perfect score: 3765
Sequence: 1 MKRRKVLPLMALSTILVSS.....LYISNPYKVYVYATKENT 735

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3519	93.5	735	1 US-08-021-601-4	Sequence 4, Appl1
2	3519	93.5	735	1 US-08-082-849B-4	Sequence 4, Appl1
3	3519	93.5	735	5 PCT-US94-01624-4	Sequence 4, Appl1
4	3519	93.5	903	1 US-08-021-601-12	Sequence 12, Appl1
5	3519	93.5	903	1 US-08-082-849B-12	Sequence 12, Appl1
6	3519	93.5	903	5 PCT-US94-01624-12	Sequence 12, Appl1
7	3342	88.8	719	1 US-08-082-849B-31	Sequence 31, Appl1
8	3342	88.8	719	5 PCT-US94-01624-31	Sequence 31, Appl1
9	783.5	20.8	884	2 US-08-471-033-5	Sequence 5, Appl1
10	783.5	20.8	884	2 US-08-471-044-5	Sequence 5, Appl1
11	783.5	20.8	884	2 US-08-463-483A-5	Sequence 5, Appl1
12	783.5	20.8	884	2 US-08-471-046A-5	Sequence 5, Appl1
13	783.5	20.8	884	2 US-08-470-566B-5	Sequence 5, Appl1
14	783.5	20.8	884	2 US-08-469-334-5	Sequence 5, Appl1
15	783.5	20.8	884	3 US-09-300-529-5	Sequence 5, Appl1
16	783.5	20.8	1346	1 US-08-471-033-23	Sequence 23, Appl1
17	783.5	20.8	1346	2 US-08-471-044-23	Sequence 23, Appl1
18	783.5	20.8	1346	2 US-08-463-483A-23	Sequence 23, Appl1
19	783.5	20.8	1346	2 US-08-471-046A-23	Sequence 23, Appl1
20	783.5	20.8	1346	2 US-08-470-566B-23	Sequence 23, Appl1
21	783.5	20.8	1346	2 US-08-469-334-23	Sequence 23, Appl1
22	783.5	20.8	1346	3 US-09-300-529-23	Sequence 23, Appl1
23	781	20.7	881	4 US-08-960-780-32	Sequence 32, Appl1
24	781	20.7	881	4 US-09-073-898-32	Sequence 32, Appl1
25	776.5	20.6	784	4 US-09-371-913A-7	Sequence 7, Appl1
26	771	20.5	1338	1 US-08-471-033-50	Sequence 50, Appl1
27	771	20.5	1338	2 US-08-471-044-50	Sequence 50, Appl1

28	771	20.5	1338	2 US-08-463-483A-50	Sequence 50, Appl1
29	771	20.5	1338	2 US-08-471-046A-50	Sequence 50, Appl1
30	771	20.5	1338	2 US-08-470-566B-50	Sequence 50, Appl1
31	771	20.5	1338	2 US-08-469-334-50	Sequence 50, Appl1
32	771	20.5	1338	3 US-09-300-529-50	Sequence 50, Appl1
33	769	20.4	852	1 US-08-471-033-36	Sequence 36, Appl1
34	769	20.4	852	2 US-08-471-044-36	Sequence 36, Appl1
35	769	20.4	852	2 US-08-463-483A-36	Sequence 36, Appl1
36	769	20.4	852	2 US-08-471-046A-36	Sequence 36, Appl1
37	769	20.4	852	2 US-08-470-566B-36	Sequence 36, Appl1
38	769	20.4	852	2 US-08-469-334-36	Sequence 36, Appl1
39	769	20.4	852	3 US-09-300-529-36	Sequence 36, Appl1
40	755.5	20.1	834	1 US-08-471-033-21	Sequence 21, Appl1
41	755.5	20.1	834	2 US-08-471-044-21	Sequence 21, Appl1
42	755.5	20.1	834	2 US-08-463-483A-21	Sequence 21, Appl1
43	755.5	20.1	834	2 US-08-471-046A-21	Sequence 21, Appl1
44	755.5	20.1	834	2 US-08-470-566B-21	Sequence 21, Appl1
45	755.5	20.1	834	2 US-08-469-334-21	Sequence 21, Appl1

ALIGNMENTS

RESULT 1
US-08-021-601-4
Sequence 4, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414,057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
FAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-4

Query Match 93.5% Score 3519, DB 1, Length 735:
Best Local Similarity 97.5% Pred. No. 7,1e-254;
Matches 688; Conservative 2; Mismatches 16; Indels 0; Gaps 0;
QY 30 EVKQENRLNSESSESSQGLGYPFDLNFQAPMVTSTGDSLPSSELENIPSENGYF 89

```
Db 1 EVKQENRLNSESSESSQGLGYYFSDNLOAPMVYTSSTTGDLSIPSSLENIIPSENOYF 60
QY 90 QSAIWSGFIKVKKSDYEYFATSADNHVTMWVDDQEVINKASNSNKRILEKGRLYOIKIOY 149
Db 61 QSAIWSGFIKVKKSDYEYFATSADNHVTMWVDDQEVINKASNSNKRILEKGRLYOIKIOY 120
QY 150 QENPTEKLDKRLVTSQNKKEVYISSDNLQPELKOKSSNRKRKRSISAGPYPPDRN 209
Db 121 QENPTEKLDKRLVTSQNKKEVYISSDNLQPELKOKSSNRKRKRSISAGPYPPDRN 180
QY 210 DSIPLSLEVEGYTVGVKKNRFLSPWISNIHEKKGITKXSSPEKMSASDYSDEKVT 269
Db 181 DSIPLSLEVEGYTVGVKKNRFLSPWISNIHEKKGITKXSSPEKMSASDYSDEKVT 240
QY 270 GRIDKNVSEARHPVLAAYPIVHVDMENTIISKNEQOSTONTDSETRTSKNTSRHT 329
Db 241 GRIDKNVSEARHPVLAAYPIVHVDMENTIISKNEQOSTONTDSETRTSKNTSRHT 300
QY 330 SEVHGNAEVHANTSRRTTSEVHGNAEVHAYADHSLSLAERTWAETMGLNTADTARL 389
Db 301 SEVHGNAEVHANTSRRTTSEVHGNAEVHAYADHSLSLAERTWAETMGLNTADTARL 360
QY 390 NANIRYVNTGTAPIYVNLPTTSLVVGKNOTLATIKAKENQLSQILAPNNYYSKNLAPIA 449
Db 361 NANIRYVNTGTAPIYVNLPTTSLVVGKNOTLATIKAKENQLSQILAPNNYYSKNLAPIA 420
QY 450 LNAODFSSPTITMNYNOFLELEKTQKRLDQYGYGNATATNFENGRRVDTGSMSEV 509
Db 421 LNAODFSSPTITMNYNOFLELEKTQKRLDQYGYGNATATNFENGRRVDTGSMSEV 480
QY 510 LPQIETTARITIFNKGDLNVERRIAANVPSDLETTRKDMTLKALKIAFGFNEPNCGL 569
Db 481 LPQIETTARITIFNKGDLNVERRIAANVPSDLETTRKDMTLKALKIAFGFNEPNCGL 540
QY 570 QYOGKDITEFDENFOOTSQNIKNQLAELNATNITYVLDKIKLNKMNILINDKRFHYDR 629
Db 541 QYOGKDITEFDENFOOTSQNIKNQLAELNATNITYVLDKIKLNKMNILINDKRFHYDR 600
QY 630 NNIAVGADESIVKKAHREYINSTEGLLINDKDIRKILSGYIVETEDTEGLKEVIYNDRY 689
Db 601 NNIAVGADESIVKKAHREYINSTEGLLINDKDIRKILSGYIVETEDTEGLKEVIYNDRY 660
QY 661 DMLNTSSLRQDGKTFIDFKKYNDKLPYISNPKNYVAVYAKENT 706
Db 661 DMLNTSSLRQDGKTFIDFKKYNDKLPYISNPKNYVAVYAKENT 706

RESULT 2
US-08-082-849B-4
: Sequence 4, Application US/08082849B
: Patent No. 5677274
: GENERAL INFORMATION:
: APPLICANT: Lepola, Stephen H.
: APPLICANT: Klompel, Kurt R.
: APPLICANT: Arota, Naveen
: APPLICANT: Singh, Yogendra
: APPLICANT: Nichols, Peter J.
: TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
: NUMBER OF SEQUENCES: 35
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/082,849B
: FILING DATE: 25-JUN-1993
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/021,601
: FILING DATE: 12-FEB-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Weber, Kenneth A.
: REGISTRATION NUMBER: 31,677
: REFERENCE/DOCKET NUMBER: 15280-161-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 735 amino acids
: type: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-082-849B-4
```

```
Query Match 93.5%: Score 3519; DB 1; Length 735;
Best Local Similarity 97.5%: Pred. No. 7,1e-234;
Matches 688; Conservative 2; Mismatches 16; Indels 0; Gaps 0;
```

```
QY 30 EVKQENRLNSESSESSQGLGYYFSDNLOAPMVYTSSTTGDLSIPSSLENIIPSENOYF 89
Db 1 EVKQENRLNSESSESSQGLGYYFSDNLOAPMVYTSSTTGDLSIPSSLENIIPSENOYF 60
QY 90 QSAIWSGFIKVKKSDYEYFATSADNHVTMWVDDQEVINKASNSNKRILEKGRLYOIKIOY 149
Db 61 QSAIWSGFIKVKKSDYEYFATSADNHVTMWVDDQEVINKASNSNKRILEKGRLYOIKIOY 120
QY 150 QENPTEKLDKRLVTSQNKKEVYISSDNLQPELKOKSSNRKRKRSISAGPYPPDRN 209
Db 121 QENPTEKLDKRLVTSQNKKEVYISSDNLQPELKOKSSNRKRKRSISAGPYPPDRN 180
QY 210 DSIPLSLEVEGYTVGVKKNRFLSPWISNIHEKKGITKXSSPEKMSASDYSDEKVT 269
Db 181 DSIPLSLEVEGYTVGVKKNRFLSPWISNIHEKKGITKXSSPEKMSASDYSDEKVT 240
QY 270 GRIDKNVSEARHPVLAAYPIVHVDMENTIISKNEQOSTONTDSETRTSKNTSRHT 329
Db 241 GRIDKNVSEARHPVLAAYPIVHVDMENTIISKNEQOSTONTDSETRTSKNTSRHT 300
QY 330 SEVHGNAEVHANTSRRTTSEVHGNAEVHAYADHSLSLAERTWAETMGLNTADTARL 389
Db 301 SEVHGNAEVHANTSRRTTSEVHGNAEVHAYADHSLSLAERTWAETMGLNTADTARL 360
QY 390 NANIRYVNTGTAPIYVNLPTTSLVVGKNOTLATIKAKENQLSQILAPNNYYSKNLAPIA 449
Db 361 NANIRYVNTGTAPIYVNLPTTSLVVGKNOTLATIKAKENQLSQILAPNNYYSKNLAPIA 420
QY 450 LNAODFSSPTITMNYNOFLELEKTQKRLDQYGYGNATATNFENGRRVDTGSMSEV 509
Db 421 LNAODFSSPTITMNYNOFLELEKTQKRLDQYGYGNATATNFENGRRVDTGSMSEV 480
QY 510 LPQIETTARITIFNKGDLNVERRIAANVPSDLETTRKDMTLKALKIAFGFNEPNCGL 569
Db 481 LPQIETTARITIFNKGDLNVERRIAANVPSDLETTRKDMTLKALKIAFGFNEPNCGL 540
QY 570 QYOGKDITEFDENFOOTSQNIKNQLAELNATNITYVLDKIKLNKMNILINDKRFHYDR 629
Db 541 QYOGKDITEFDENFOOTSQNIKNQLAELNATNITYVLDKIKLNKMNILINDKRFHYDR 600
QY 630 NNIAVGADESIVKKAHREYINSTEGLLINDKDIRKILSGYIVETEDTEGLKEVIYNDRY 689
Db 601 NNIAVGADESIVKKAHREYINSTEGLLINDKDIRKILSGYIVETEDTEGLKEVIYNDRY 660
QY 660 DMLNTSSLRQDGKTFIDFKKYNDKLPYISNPKNYVAVYAKENT 735
Db 660 DMLNTSSLRQDGKTFIDFKKYNDKLPYISNPKNYVAVYAKENT 735
```

Db 661 DMLNSSLRODGKTFIDFKKYNKDLPLYISNPYKVVAVTKENT 706

RESULT 3

PCT-US94-01624-4

Sequence 4, Application PC/TUS9401624

GENERAL INFORMATION:

APPLICANT: Leppia, Stephen H.

APPLICANT: Kilmpel, Kurt R.

APPLICANT: Atora, Naveen

APPLICANT: Singh, Yogendra

APPLICANT: Nichols, Peter J.

TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND

TITLE OF INVENTION: RELATED METHODS

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW

STREET: Stewart Street Tower, 20th Floor, One Market

STREET: Plaza

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94105

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/01624

FILING DATE: June 25, 1993

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Weber, Kenneth A.

REGISTRATION NUMBER: 31,677

REFERENCE/DOCKET NUMBER: 15280-115

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 543-9600

TELEFAX: (415) 543-5043

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 735 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US94-01624-4

Query Match 93.5%; Score 3519; DB 5; Length 735;

Best Local Similarity 97.5%; Pred. No. 7, 1e-254;

Matches 688; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 30 EYKQENRLNSESSESSQGLIGYFSDLNFOAPNVVTSSTTGDISPSSSELENIPSENQYF 89

Db 1 EYKQENRLNSESSESSQGLIGYFSDLNFOAPNVVTSSTTGDISPSSSELENIPSENQYF 60

QY 90 QSAIWSGFTIKVKSDEYTPATSDNHNVTMVDQEVYINKSNSNKRIRLENGRLYQIKIY 149

Db 61 QSAIWSGFTIKVKSDEYTPATSDNHNVTMVDQEVYINKSNSNKRIRLENGRLYQIKIY 120

QY 150 QRENPEKGLDFKLYMTDSQNKKEVYSSDNLQJPELKQKSSNRKRKSTAGPTVPDRDN 209

Db 121 QRENPEKGLDFKLYMTDSQNKKEVYSSDNLQJPELKQKSSNRKRKSTAGPTVPDRDN 180

QY 210 DGIPLDSLEVEGYTVDVKNKRTFLSPMISNHEKGLTKYSSPEKSTASDPYSDPEKVT 269

Db 181 DGIPLDSLEVEGYTVDVKNKRTFLSPMISNHEKGLTKYSSPEKSTASDPYSDPEKVT 240

QY 270 GRIDKVVSPARHPPLVAAPIYIVVDMENTILSKNEDOSTONTSETRTISKNTSTSRTHT 329

Db 241 GRIDKVVSPARHPPLVAAPIYIVVDMENTILSKNEDOSTONTSETRTISKNTSTSRTHT 300

QY 330 SEVHGNAEVHANTSTRTHTSEVHGNAEVHAAVAIDHSLSLAGERTMAETGLNTADTARL 389

Db 301 SEVHGNAEVHASFEDIGGSVSAGFSNNSSTVAIDHSLSLAGERTMAETGLNTADTARL 360

QY 390 NANIRVYNTGTAPLYNLPPTTSLVLGKNQTLATIKAKENOLSOILAPNNYPSKNLAPIA 449

Db 361 NANIRVYNTGTAPLYNLPPTTSLVLGKNQTLATIKAKENOLSOILAPNNYPSKNLAPIA 420

QY 450 LNAODPESSPTITMNNVQFLEKTKQLRLDTQOVYNITNTYFENGRVAVDGSNNSEV 509

Db 421 LNAODPESSPTITMNNVQFLEKTKQLRLDTQOVYNITNTYFENGRVAVDGSNNSEV 480

QY 510 LPQIETARTIIFNGKDLNVERRIAANVPSDETTKPMTLKEALKIAFGNEPAGNL 569

Db 481 LPQIETARTIIFNGKDLNVERRIAANVPSDETTKPMTLKEALKIAFGNEPAGNL 540

QY 570 QYQKDTIEFDENFDQTSQNIKNQLAELNATNLYVLDKIKLNKKNLIRDKRFHYDR 629

Db 541 QYQKDTIEFDENFDQTSQNIKNQLAELNATNLYVLDKIKLNKKNLIRDKRFHYDR 600

QY 630 NNIAVGADESIVKEAHEVINSSTEGLLNIDDKIRKILSGYIEIEDTGLKEVINDRY 689

Db 601 NNIAVGADESIVKEAHEVINSSTEGLLNIDDKIRKILSGYIEIEDTGLKEVINDRY 660

QY 690 DMLNSSLRODGKTFIDFKKYNKDLPLYISNPYKVVAVTKENT 735

Db 661 DMLNSSLRODGKTFIDFKKYNKDLPLYISNPYKVVAVTKENT 706

RESULT 4

US-08-021-601-12

Sequence 12, Application US/08021601

Patent No. 5591631

GENERAL INFORMATION:

APPLICANT: Leppia, Stephen H.

APPLICANT: Kilmpel, Kurt R.

APPLICANT: Nichols, Peter J.

APPLICANT: Atora, Naveen

APPLICANT: Singh, Yogendra

TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND

TITLE OF INVENTION: RELATED METHODS

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Needle & Rosenberg, P.C.

STREET: 133 Carnegie Way, Suite 400

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/021,601

FILING DATE: 19930212

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Spratl, Gwendolyn D.

REGISTRATION NUMBER: 36,016

REFERENCE/DOCKET NUMBER: 1414,057

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404/688-0770

TELEFAX: 404/688-9880

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 903 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-021-601-12

Query Match 93.5%; Score 3519; DB 1; Length 903;
Best Local Similarity 97.5%; Pred. No. 9.7e-254;
Matches 688; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

30 EVKQENRLNSESSESSOGGLGYYFSDLNFOAPMVYTSSTGDLSPSSLENIPESENOYF 89
1 EVKQENRLNSESSESSOGGLGYYFSDLNFOAPMVYTSSTGDLSPSSLENIPESENOYF 60
90 OSATSGFITYKKSDEYTFATSDAHNVYMWDDQEVINKASNSKIRLEKGLYQIKIY 149
61 OSATSGFITYKKSDEYTFATSDAHNVYMWDDQEVINKASNSKIRLEKGLYQIKIY 120
150 QRENTEGGLDKLWYTSQNKKEVYSSDNLOLPELKOKSSNRKRKSTASAPYPPDRN 209
121 QRENTEGGLDKLWYTSQNKKEVYSSDNLOLPELKOKSSNRKRKSTASAPYPPDRN 180
210 DGIPLSLEVEGYTVVKNKRFPFLSPWISNIEHKGLTYKSSPEKMSASDPYSDFEKYT 269
181 DGIPLSLEVEGYTVVKNKRFPFLSPWISNIEHKGLTYKSSPEKMSASDPYSDFEKYT 240
270 GRIDKNVSPERHPLVAAYPIVHVDMENIILSKNDOSTQNTDSTRTISKNTSRTHT 329
241 GRIDKNVSPERHPLVAAYPIVHVDMENIILSKNDOSTQNTDSTRTISKNTSRTHT 300
330 SEVHGNAEYHANTSTRHTSEVHGNAEYHAYADHSLSLAGEERTWAGTGLNTADTARL 389
301 SEVHGNAEYHANTSTRHTSEVHGNAEYHAYADHSLSLAGEERTWAGTGLNTADTARL 360
390 NANIRYVNTGTAPIYVNLPTTSLVIGKNOTLATIRAKENQSLQILAPNNYPSKMLADIA 449
361 NANIRYVNTGTAPIYVNLPTTSLVIGKNOTLATIRAKENQSLQILAPNNYPSKMLADIA 420
450 LNAODDESSPTITMNYNOFLELEKTKQLRLTDQYGYGNATYNFENGVRVDTGSMSEV 509
421 LNAODDESSPTITMNYNOFLELEKTKQLRLTDQYGYGNATYNFENGVRVDTGSMSEV 480
510 LPOIETTARIIFNGKDLNVERRIAANVPSDPLETTKPDMTIKALKIAGFNPNGNL 569
481 LPOIETTARIIFNGKDLNVERRIAANVPSDPLETTKPDMTIKALKIAGFNPNGNL 540
570 QYQGDITEFEPNFDQOSTONIKNOELAMATNITYVLDKIKLNKMMILLRDKRFHNDY 629
541 QYQGDITEFEPNFDQOSTONIKNOELAMATNITYVLDKIKLNKMMILLRDKRFHNDY 600
630 NNIAVGADESVYKKAHREYINSSTEGLLNIDKDIRKILSGYIEIEDTEGLKEVINDRY 689
601 NNIAVGADESVYKKAHREYINSSTEGLLNIDKDIRKILSGYIEIEDTEGLKEVINDRY 660
690 DMLNSSLRQDGKTFIDPKKYNKDLPLIYISNPYKVNYAYATKENT 735
661 DMLNSSLRQDGKTFIDPKKYNKDLPLIYISNPYKVNYAYATKENT 706

RESULT 5
US-08-082-849B-12
Sequence 12, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Kimpel, Kurt R.
APPLICANT: Arora, Navene
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
TITLE OF INVENTION: Related Methods
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-12

Query Match 93.5%; Score 3519; DB 1; Length 903;
Best Local Similarity 97.5%; Pred. No. 9.7e-254;
Matches 688; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

30 EVKQENRLNSESSESSOGGLGYYFSDLNFOAPMVYTSSTGDLSPSSLENIPESENOYF 89
1 EVKQENRLNSESSESSOGGLGYYFSDLNFOAPMVYTSSTGDLSPSSLENIPESENOYF 60
90 OSATSGFITYKKSDEYTFATSDAHNVYMWDDQEVINKASNSKIRLEKGLYQIKIY 149
61 OSATSGFITYKKSDEYTFATSDAHNVYMWDDQEVINKASNSKIRLEKGLYQIKIY 120
150 QRENTEGGLDKLWYTSQNKKEVYSSDNLOLPELKOKSSNRKRKSTASAPYPPDRN 209
121 QRENTEGGLDKLWYTSQNKKEVYSSDNLOLPELKOKSSNRKRKSTASAPYPPDRN 180
210 DGIPLSLEVEGYTVVKNKRFPFLSPWISNIEHKGLTYKSSPEKMSASDPYSDFEKYT 269
181 DGIPLSLEVEGYTVVKNKRFPFLSPWISNIEHKGLTYKSSPEKMSASDPYSDFEKYT 240
270 GRIDKNVSPERHPLVAAYPIVHVDMENIILSKNDOSTQNTDSTRTISKNTSRTHT 329
241 GRIDKNVSPERHPLVAAYPIVHVDMENIILSKNDOSTQNTDSTRTISKNTSRTHT 300
330 SEVHGNAEYHANTSTRHTSEVHGNAEYHAYADHSLSLAGEERTWAGTGLNTADTARL 389
301 SEVHGNAEYHANTSTRHTSEVHGNAEYHAYADHSLSLAGEERTWAGTGLNTADTARL 360
390 NANIRYVNTGTAPIYVNLPTTSLVIGKNOTLATIRAKENQSLQILAPNNYPSKMLADIA 449
361 NANIRYVNTGTAPIYVNLPTTSLVIGKNOTLATIRAKENQSLQILAPNNYPSKMLADIA 420
450 LNAODDESSPTITMNYNOFLELEKTKQLRLTDQYGYGNATYNFENGVRVDTGSMSEV 509
421 LNAODDESSPTITMNYNOFLELEKTKQLRLTDQYGYGNATYNFENGVRVDTGSMSEV 480
510 LPOIETTARIIFNGKDLNVERRIAANVPSDPLETTKPDMTIKALKIAGFNPNGNL 569
481 LPOIETTARIIFNGKDLNVERRIAANVPSDPLETTKPDMTIKALKIAGFNPNGNL 540
570 QYQGDITEFEPNFDQOSTONIKNOELAMATNITYVLDKIKLNKMMILLRDKRFHNDY 629
541 QYQGDITEFEPNFDQOSTONIKNOELAMATNITYVLDKIKLNKMMILLRDKRFHNDY 600
630 NNIAVGADESVYKKAHREYINSSTEGLLNIDKDIRKILSGYIEIEDTEGLKEVINDRY 689

Db 601 NNIAVGADESVEKAEHREVINSSTEGLLNIDKDIRKILSGYIIEIDTEGLKEVINDRY 660
QY 690 DMLNISSLRODGKTFIDFKKRYNDKPLIYSNPYKVVAVTKENT 735
Db 661 DMLNISSLRODGKTFIDFKKRYNDKPLIYSNPYKVVAVTKENT 706

RESULT 6

PCT-US94-01624-12
Sequence 12, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
STREET: Stewart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ. ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-12

Query Match 93.5%; Score 3519; DB 5; Length 903;
Best Local Similarity 97.5%; Pred. No. 9, 7e-254;
Matches 688; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 30 EYVQENRLNESSESSGGLIGYFSDINFOAPWVYISSTTGDISPSELENIPESEQYF 89
Db 1 EYVQENRLNESSESSGGLIGYFSDINFOAPWVYISSTTGDISPSELENIPESEQYF 60
QY 90 QSAIWSGFIKVKSSDETFATSDNHTMVDVDOEVINKASNSKIRLEKGRLOYQI 149
Db 61 QSAIWSGFIKVKSSDETFATSDNHTMVDVDOEVINKASNSKIRLEKGRLOYQI 120
QY 150 QRENPTKGLDFKLYTDSQNKKEVSSDNLOLPELKQKSSNSRKRSTAGPTVPDRDN 209
Db 121 QRENPTKGLDFKLYTDSQNKKEVSSDNLOLPELKQKSSNSRKRSTAGPTVPDRDN 180
QY 210 DGIPDSLEVEGYVVDVKNKFTFLSPWISNIEKKGLTKRYSSPEKWSASDPYSDFE 269
Db 181 DGIPDSLEVEGYVVDVKNKFTFLSPWISNIEKKGLTKRYSSPEKWSASDPYSDFE 240

QY 270 GRIDKNVSPARRPLVAAPYIVHVDMENILISKNEOSTONTDSEFTTISKNTSRTHT 329
Db 241 GRIDKNVSPARRPLVAAPYIVHVDMENILISKNEOSTONTDSEFTTISKNTSRTHT 300
QY 330 SEVHGNAEVHANTSTSTHTSEVHGNAEVAHAVIDHSLSLAGEWTAGLTADTARL 389
Db 301 SEVHGNAEVHANTSTSTHTSEVHGNAEVAHAVIDHSLSLAGEWTAGLTADTARL 360
QY 390 NANIRYVNTGAPLYNLPFTSLYLGNQTLATIKAKENOLSOILAPNNYPPSKNLAPIA 449
Db 361 NANIRYVNTGAPLYNLPFTSLYLGNQTLATIKAKENOLSOILAPNNYPPSKNLAPIA 420
QY 450 LNAODPSSPTITMNYNQFIELEKTKOLRDTDOVGNITATYFENGRVAVDGSNNSEV 509
Db 421 LNAODPSSPTITMNYNQFIELEKTKOLRDTDOVGNITATYFENGRVAVDGSNNSEV 480
QY 510 LPOIETTARIIFNGKDLNVERRIAANVNSDPLETKPMTLEKALIAFGNEPNCNL 569
Db 481 LPOIETTARIIFNGKDLNVERRIAANVNSDPLETKPMTLEKALIAFGNEPNCNL 540
QY 570 QYOGKITEFDNFNDQOSTONIKNOELATNITYVLKIKLAKANNILIRDKRFHYDR 629
Db 541 QYOGKITEFDNFNDQOSTONIKNOELATNITYVLKIKLAKANNILIRDKRFHYDR 600
QY 630 NNIAVGADESVEKAEHREVINSSTEGLLNIDKDIRKILSGYIIEIDTEGLKEVINDRY 689
Db 601 NNIAVGADESVEKAEHREVINSSTEGLLNIDKDIRKILSGYIIEIDTEGLKEVINDRY 660
QY 690 DMLNISSLRODGKTFIDFKKRYNDKPLIYSNPYKVVAVTKENT 735
Db 661 DMLNISSLRODGKTFIDFKKRYNDKPLIYSNPYKVVAVTKENT 706

RESULT 7

US-08-082-849B-31
Sequence 31, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ. ID NO: 31:

SEQUENCE CHARACTERISTICS:
LENGTH: 719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
us-08-082-8498-31

Query Match 88.8%; Score 3342; DB 1; Length 719;
Best Local Similarity 92.8%; Pred. No. 1e-240;
Matches 659; Conservative 6; Mismatches 21; Indels 24; Gaps 2;

```

OY 30 EYKQENRLINESSSOGGLGYFSDLNQAPMVYTSSTGDLSPSSLENIPESENOYF 89
    |||||||
DB 1 EYKQENRLINESSSOGGLGYFSDLNQAPMVYTSSTGDLSPSSLENIPESENOYF 60
OY 90 OSAIWSGFIKYKKSDEYFATSDADNHVMMVDDQEVINKASNSNRILEKGLYQIKIY 149
    |||||||
DB 61 OSAIWSGFIKYKKSDEYFATSDADNHVMMVDDQEVINKASNSNRILEKGLYQIKIY 120
OY 150 QRENPTKGLDFKLYWTDSONKKEVISSDNQLPELKOKSSNTATIMQKGFLOGPTVP 205
    |||||||
DB 121 QRENPTKGLDFKLYWTDSONKKEVISSDNQLPELKOKSSNTATIMQKGFLOGPTVP 180
OY 206 DRDNGCIDDSLEVEGYTVDVKNKRFTLSPWISNIHEKKGLTYKSSPEKWSASDPYSDF 265
    |||||||
DB 181 DRDNGCIDDSLEVEGYTVDVKNKRFTLSPWISNIHEKKGLTYKSSPEKWSASDPYSDF 240
OY 266 EKVYGRIDKNVSPKARHPLVAAYPIVHVDMEIILSKNEDOSTQMTDETRISKNTSTS 325
    |||||||
DB 241 EKVYGRIDKNVSPKARHPLVAAYPIVHVDMEIILSKNEDOSTQMTDETRISKNTSTS 300
OY 326 RTHTSEVHGNAEVAHANTSTSRTHTSEVHGNAEVAHVAIDHSLSLAGERTWAEIATGLNTAD 385
    |||||||
DB 301 RTHTSEVHGNAEVAHANTSTSRTHTSEVHGNAEVAHVAIDHSLSLAGERTWAEIATGLNTAD 360
OY 386 TARLANANIRYVNTGTAPLYNVLPTTSLVIGKNOTLATIKAKENQSLQILAPNNYPSKNL 445
    |||||||
DB 361 TARLANANIRYVNTGTAPLYNVLPTTSLVIGKNOTLATIKAKENQSLQILAPNNYPSKNL 420
OY 446 APIALNAODDFSSPTITMNYNOFLELEKTKQLRLDQVYGNIAIYNNENGRVAVDTGSN 505
    |||||||
DB 421 APIALNAODDFSSPTITMNYNOFLELEKTKQLRLDQVYGNIAIYNNENGRVAVDTGSN 460
OY 506 WSEVLPQIOETTARLIIFNGKDLNLVERRIAANVPSDPLETTKPDMTLKEALKIAGFNEP 565
    |||||||
DB 461 WSEVLPQIOETTARLIIFNGKDLNLVERRIAANVPSDPLETTKPDMTLKEALKIAGFNEP 520
OY 566 NGNLOYGKDIETEPFNFDQOTSONIKQOLAEIATNTIYTVLDKIKLAKMMILLRDKRF 625
    |||||||
DB 521 NGNLOYGKDIETEPFNFDQOTSONIKQOLAEIATNTIYTVLDKIKLAKMMILLRDKRF 580
OY 626 HYDRNNIIVAGDESYYKKAHREVIINSTEGLLINDIKDIRILSGIYIEIDTEGLKEVI 685
    |||||||
DB 581 HYDRNNIIVAGDESYYKKAHREVIINSTEGLLINDIKDIRILSGIYIEIDTEGLKEVI 640
OY 686 NDRYDMLNISSLRDQKTFIDFKKYNDKLPYISNPNYKVVNYAVATKENT 735
    |||||||
DB 641 NDRYDMLNISSLRDQKTFIDFKKYNDKLPYISNPNYKVVNYAVATKENT 690

RESULT 8
PCT-US94-01624-31
; Sequence 31, Application PC/TUS9401624
; GENERAL INFORMATION:
; APPLICANT: Leppa, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Arota, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 31
```

```

; CORRESPONDENCE ADDRESS:
; ADDRESS: TOWNSEND and TOWNSEND KHOURIE and CREW
; STREET: Stewart Street Tower, 20th Floor, One Market
; STREET: Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01624
; FILING DATE: June 25, 1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-115
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-01624-31
```

```

Query Match 88.8%; Score 3342; DB 5; Length 719;
Best Local Similarity 92.8%; Pred. No. 1e-240;
Matches 659; Conservative 6; Mismatches 21; Indels 24; Gaps 2;

OY 30 EYKQENRLINESSSOGGLGYFSDLNQAPMVYTSSTGDLSPSSLENIPESENOYF 89
    |||||||
DB 1 EYKQENRLINESSSOGGLGYFSDLNQAPMVYTSSTGDLSPSSLENIPESENOYF 60
OY 90 OSAIWSGFIKYKKSDEYFATSDADNHVMMVDDQEVINKASNSNRILEKGLYQIKIY 149
    |||||||
DB 61 OSAIWSGFIKYKKSDEYFATSDADNHVMMVDDQEVINKASNSNRILEKGLYQIKIY 120
OY 150 QRENPTKGLDFKLYWTDSONKKEVISSDNQLPELKOKSSNTATIMQKGFLOGPTVP 205
    |||||||
DB 121 QRENPTKGLDFKLYWTDSONKKEVISSDNQLPELKOKSSNTATIMQKGFLOGPTVP 180
OY 206 DRDNGCIDDSLEVEGYTVDVKNKRFTLSPWISNIHEKKGLTYKSSPEKWSASDPYSDF 265
    |||||||
DB 181 DRDNGCIDDSLEVEGYTVDVKNKRFTLSPWISNIHEKKGLTYKSSPEKWSASDPYSDF 240
OY 266 EKVYGRIDKNVSPKARHPLVAAYPIVHVDMEIILSKNEDOSTQMTDETRISKNTSTS 325
    |||||||
DB 241 EKVYGRIDKNVSPKARHPLVAAYPIVHVDMEIILSKNEDOSTQMTDETRISKNTSTS 300
OY 326 RTHTSEVHGNAEVAHANTSTSRTHTSEVHGNAEVAHVAIDHSLSLAGERTWAEIATGLNTAD 385
    |||||||
DB 301 RTHTSEVHGNAEVAHANTSTSRTHTSEVHGNAEVAHVAIDHSLSLAGERTWAEIATGLNTAD 360
OY 386 TARLANANIRYVNTGTAPLYNVLPTTSLVIGKNOTLATIKAKENQSLQILAPNNYPSKNL 445
    |||||||
DB 361 TARLANANIRYVNTGTAPLYNVLPTTSLVIGKNOTLATIKAKENQSLQILAPNNYPSKNL 420
OY 446 APIALNAODDFSSPTITMNYNOFLELEKTKQLRLDQVYGNIAIYNNENGRVAVDTGSN 505
    |||||||
DB 421 APIALNAODDFSSPTITMNYNOFLELEKTKQLRLDQVYGNIAIYNNENGRVAVDTGSN 460
OY 506 WSEVLPQIOETTARLIIFNGKDLNLVERRIAANVPSDPLETTKPDMTLKEALKIAGFNEP 565
    |||||||
DB 461 WSEVLPQIOETTARLIIFNGKDLNLVERRIAANVPSDPLETTKPDMTLKEALKIAGFNEP 520
```



```

1 NUMBER OF SEQUENCES: 50
2 CORRESPONDENCE ADDRESS:
3 ADDRESSEE: CIBA-GEIGY Corporation
4 STREET: 7 Skyline Drive
5 CITY: Hawthorne
6 STATE: NY
7 COUNTRY: USA
8 ZIP: 10532
9
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: Floppy disk
12 COMPUTER: IBM PC compatible
13 OPERATING SYSTEM: PC-DOS/MS-DOS
14 SOFTWARE: PatentIn Release #1.0, Version #1.30B
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/08/471,044
17 FILING DATE: 06-JUN-1995
18 CLASSIFICATION: 800
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER: US 08/463,483
21 FILING DATE: 05-JUN-1995
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: US 08/314,594
24 FILING DATE: 09-SEP-1994
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: US 08/218,018
27 FILING DATE: 23-MAR-1994
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: US 08/037,057
30 FILING DATE: 25-MAR-1993
31 ATTORNEY/AGENT INFORMATION:
32 NAME: Pace, Gary M.
33 REGISTRATION NUMBER: 40,403
34 REFERENCE/DOCKET NUMBER: 40,403
35 TELECOMMUNICATION INFORMATION:
36 TELEPHONE: 919-541-8582
37 TELEFAX: 919-541-8689
38 INFORMATION FOR SEQ ID NO: 5:
39 SEQUENCE CHARACTERISTICS:
40 LENGTH: 884 amino acids
41 TYPE: amino acid
42 TOPOLOGY: linear
43 MOLECULE TYPE: protein
44
45 US-08-471-044-5

```

Query Match	20.8%	Score 783.5	DB 2	Length 884
Best Local Similarity	29.6%	Pred. 4.2e-50		
Matches 234	Conservative 133	Mismatches 293	Indels 131	Gaps 29

QY	1	MKKRRVLLPLMALSTILVSS--TGNLEVIOAEVK-----QENRLNLESSSOGILGY	51
		:	
Db	1	MKNMKKTLASVYTCITLAPMFLNGVNVNAVYADSKTNQISTTQKNQ--QKENDRKGILGY	57
		:	
QY	52	YFSDLNFCAPMNVYVTSSTGDLSPSELEN--IPSENOYFGSAIISGFIKAVKSDBYTFA	109
		:	
Db	58	YFKGKDF-SNLTWEPTRDSTLIYDQATANKLLDKKOEOYOSIRWIGLQSKETGDFTFN	116
		:	
QY	110	TSADHNHVMVDDQEVIMKNASNSNKIRLEKRGLYOIKIOYENPTEKLD-----FKL	163
		:	
Db	117	LSEDQALIEINGKLIISNKGKQYVHLEKGLVPLKIEYQSD--TKFNIDSKTRKELK	174
		:	
QY	164	YWTDSQNKREVISSDNLQLPKLKQSS-----NSRKRKSTSAQPTVPDROND	210
		:	
Db	175	EKIDSONOPQVOQDELNRNPFNKRESQEFIAKPSKINLFTQMKMREIDED--ITDQD	231
		:	
QY	211	GIPDSLEVEGTVDYVKNKRTFLSPVISMNHEKKGLTKIKSSPEKKNSTADPVSDEKYTG	270
		:	
Db	232	SIPDLMENGTYI-----ONRIAVKKDDSL-ASKGYTKFVSNPLSEHTQDPTDYDEKAR	286
		:	
QY	271	RIDKVVSEARPLVAAVPIYVDKNNIISNENQOSTQNTDSERTITSKNSTSRTHS	330
		:	
Db	287	DDUSNAKETNPVLAAPSPVAVSKEKYLSPENENIS-----	323
		:	

```

0Y 331 EVNGAEVHAHTMSPSRHTSEVHNAEVH-----AVAIHDSLAGETAEMTG-- 360Q
Db 334 ---NSVEHSSTNMSTNTEGASVEAGIGPKGISFGVSANOJSEFVAOE--WSTSTGNT 378T
0Y 381 -LNTADTARLANIRIVNNTGTAPLYVNLPTTSLVLKKNQTLATIKKENQLOILPNN 438P
Db 379 SOFNTASAGYILANRVRYNNGVTGAIYDVKPPTTSFVL--NNDTIATITAKSNTALNISPE 437T
0Y 439 YTPSKNINLAPIALNMDODSSPTPTMNTNQNLELEKTQQLRLDTDQVYGNATYVFNENGKV 439R
Db 438 SYPKGGQNGIAITSDDENSHPIITLNKKQVDNLLNNKPMLETPMOTGG---VYKIKDPTG 434T
0Y 499 RVDTSNMNSEVLPLOEFTATLIFNGKDLINVERIRIAVNPDSLETTKPDMTLKLI 558R
Db 495 NIYTGEGMNGYIQQIKATKATSIYYDQGE-RAEKRVAAKTYENEDXT-FSLTIKDLAKL 552Z
0Y 559 AF--GENEPGNLOYQGDITEPDF--NFDQOTSQNIKNQLAEL---NATNIYTVLQK 6090Y
Db 553 SYPDIEIKIEGLLYVKNKPIEYSSVMYTLIDENTAKETWKLQNDPTTQKFKFDVSHLYDV-- 6090Y
0Y 610 IKLAKNMILRLROKRFHNDRNINIVGADSESVKRAHREVINSSTREG-----LL 657Q
Db 610 -KLTPKMAVTTIK-LSILYDN---AESNDNSIGKATNTNINIVSGGNGKKOYSSNNPDANLT 664Q
0Y 658 LNID-----KDIRKILSGYIVELIETEGLEKVINDRYDMLNIS-SLRQDGKTFIDEKKY 710Q
Db 665 LNTDQOEKLNRDVIYISLYMKSEKNTQCEITIGIEIYPIPTTKTVNVNKNKDKYRLDIIAH 724Q
0Y 711 NDKLPLIYISNP 721
Db 725 NIK-----SNP 730

```

RESULT 11
 US-08-463-483A-5
 Sequence 5, Application US/08463483A
 Patent No. 5849870
 GENERAL INFORMATION:
 APPLICANT: Warren, Gregory W
 APPLICANT: Kozziel, Michael G
 APPLICANT: Mullins, Martha A
 APPLICANT: Nye, Gordon J
 APPLICANT: Carr, Brian
 APPLICANT: Desai, Nalini M
 APPLICANT: Kostelchka, N. Kristy
 APPLICANT: Duck, Nicholas B
 APPLICANT: Estruch, Juan J
 TITLE OF INVENTION: No. 5849870el Pesticidal Proteins and Strains
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: NY
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30B
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/463,483A
 FILING DATE:
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,594
 FILING DATE: 09-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/218,018
 FILING DATE: 23-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/037,057

FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-463-483A-5

Query Match 20.8%; Score 783.5; DB 2; Length 884;
Best Local Similarity 29.6%; Pred. No. 4,2e-50;
Matches 234; Conservative 133; Mismatches 293; Indels 131; Gaps 29;

QY 1 MKRKRLIPMLSTLVSS--TGNLEVIOAEVK-----QENRLNSESSESSQGLGY 51
DB 1 MKMKKKLASVVTCTLLAPMFLNGNVAAYADSKTNQISTQKNQ---QKEMDRKGLGY 57
QY 52 YFSDLNFOAPMVTSTTGDLSIPSELEN--IPSENQFQSAIWSGFIKVKSDYTF 109
DB 58 YFKGKDF-SULTMFAPTRDSTLIYDOOTANKLDDKQOEYQSTIRWIGLISKEGTFTN 116
QY 110 TSADNHVTMVDQDEVINKASNSNKRIRLEKGRLYQIKIOYORENPEKGLD----FKL 163
DB 117 LSEDEQAIIEINCKIISNKEKQVHLEKGLVPIKIEYQSD--TKFNIDSKTFKELKL 174
QY 164 YMTDSQNKKEVISDMLQELPKOKS-----NSRKRKTSAGPIVPPDND 210
DB 175 FKIDSONOPQOODELRNEEFNKKSQEFLAPRSKINLFTQMKKREIDED---TDTDGD 231
QY 211 GIPDSLEVEGYTVDKKRRFLSPWISNIHEKKGLTKYKSSPEKMSAPSDPYSPFEKVTG 270
DB 232 STIDLMENKTYT-----QNIAYKMDSL-ASKGYTFEVSNPLESIRVGDPIYDEKAAR 286
QY 271 RIDKNVSPEARHPLVAAYPIVHVMENIILSKNEDOSTONTDETRTISKNTSRTHTS 330
DB 287 DLDLSNAKETFNPLVAAFPVSNVMEKVILSPNENIS----- 323
QY 331 EVGNMNEVHANTSTSRTHSEVGNNAEVH-----AVALDLSLAGERTAEIYNG-- 380
DB 324 ---NSVESHSSTWSTYNTGASVEAGIGPKGISFGVSNVYOHSEIYAOE--WGTSTGNT 378
QY 381 --LNTADTARLNNIRYVNTGTAPIYVNLPTTSVLGKNQTLATIKAKENQLSILAPNN 438
DB 379 SORNTASAGILNNVRYNNVGTALYDKPTTSFVL--NNPTLATITAKSNSTALNISPEGE 437
QY 439 YPYSKMLAPIALNAODEFSPTTMVNOFLLEKTKQLRLDQVYGNATYFENGGRV 498
DB 438 SYPKKGONGIAITSMDENSHPTLNNKKQVDNLNNKPMLEFNQDQG---VYKIKTHG 494
QY 499 RVDTGSNMSVYLPOIQETTRITFNGKDLNVERRIAANVPSDPLETTKQDMTKELKI 558
DB 495 NIVTGGEMNGVIOIKRAKTAIIIVDGE--RVAEKRVAAKYENPEDKT--PSLTLKDALKL 552
QY 559 AF--GFNEPKNQYOOGKIDTEFDF--NPDQOTSQNIKNQLAEL-----NATNIYTVLDK 609
DB 553 SYDEDEKELIGLLYKKNKPIYESVMTYLDENTAKEVTKQNLNDTGTGKFKOVSHLYDV--- 609
QY 610 IKLNAKMLILIRDKRPHYDRNNIYAGADESVKAEHREVINSSTEG-----LL 657
DB 610 -KLTPKMNVTIK-LSTLYDN---AESNDNSIGKWTNTNINISGCGNCKKQYSSNNPDANLT 664
QY 658 LNTD-----KDIRKLISGIVIEIDTEGLKEVINDRYDMLNIS-SLRQDGKFTIDPKKY 710
DB 665 LNTDAOEKILNKNRDYYISTLYMKSEKNTQCEITIDGELIYPTTTKTVNVNKNKYKRLDITAH 724

QY 711 NDKLPLXISNP 721
DB 725 NIK-----SNP 730

RESULT 12
US-08-471-046A-5
Sequence 5, Application US/08471046A
Patent No. 5866326
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalin M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: Method for Isolating Vegetative Insecticidal
TITLE OF INVENTION: Protein Genes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5866326artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.308
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,046A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-471-046A-5

Query Match 20.8%; Score 783.5; DB 2; Length 884;
Best Local Similarity 29.6%; Pred. No. 4,2e-50;
Matches 234; Conservative 133; Mismatches 293; Indels 131; Gaps 29;

QY 1 MKRKRLIPMLSTLVSS--TGNLEVIOAEVK-----QENRLNSESSESSQGLGY 51
DB 1 MKMKKKLASVVTCTLLAPMFLNGNVAAYADSKTNQISTQKNQ---QKEMDRKGLGY 57

```

0Y 52 YFSDLNPAAPVWYSSSTGDISFSSSELEN--IPSENOYQSALWMSGIFKKSDEYTA 109
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
0Y 58 YFKGDF--SNLTPAPPTDSTLLIYDOOTANKLLDKKQOEQOSIRWIGLIDSKETGETFN 116
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
0Y 110 TSADNHVWVMDQDEVINKKASNSKRIRLEGRLYQIKIYOARENPEKGD-----FKL 163
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 117 LSEDOALIEINGKIKSKKEKQVHLEKGLVPIKIEYOSD--TFNIDSTFKRELK 174
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
0Y 164 YWTSOANKKEVYSSDNQJLEPLAKSS-----NSKKRSTSGAPVYPRDND 210
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 175 FKIDSQNPQOQVODELKNPFPENKKEQEFLLAKPSKINLFTQKKRREIDD---PTPDGD 231
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
0Y 211 GIPDSLEVEGYTVVQKMKRTFLSPWIMINIEHKKLTJYKSGSPKMSWASPYGDPFVYG .270
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 232 SIPDLMENGNTT-----QNRILAVKWDLSL-ASKETTFVSNPPEFSHTGVPYTDYEKAAR 286
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
0Y 271 RIDKNVSPEARHPVLAAPYPIVHDMENILISKNEQOSTONTDSETRTISKNTSTSRHTS 330
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 287 DLDLSNAKETENPNLVAAPFVSVMEMKYLISLPNENLS----- 323
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
0Y 331 EYHGAAEYVHANTSVSRHTTSEVHGAAEVH-----AVADHSLSLAGERTAEIWMG-- 380
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 334 ---NSVESHSSTMSVNTTEGASVAGIGPKGISFGVSVMYOHSEYIAOE--WGTSTGNT 378
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
0Y 381 --LNTADARLANANIRVYNTGTAPYVNLPTTSLVLGKNOTLATIRKAKENQLSOILAPN 438
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 379 SQFNWASGYLANAVRYNNVGTGALIVYKPTTSVL--NDTIATITAKSSTLAINISPGE 437
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
0Y 439 YPSPKNLAPIALNADDEFSPTPTMANYNQPLELEKTQOLRDLTDQVYGNATYATNFENGRV 498
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 438 SYPKKGQNGIATISMDFFNSHPILTNKKQYDNLNKNPKMLLETJNQTFDG---VYKIDTGG 494
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
0Y 499 RYDGSNMSEVLPOIOETTAIIINGDMLYVERIRIAYVNSPDETTKRDMLTKALKI 558
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 495 NIWVGEGNNGYIQOIKAKTASTIYDDGE-RVAEKRYAAKDYENEDKT--PSLLTKALXKI 552
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
0Y 559 AF--GFNEPNCNLOYOGKDITPEPF--NFDQOSTQNIKNOLAEL-----NATNITYVLDK 609
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 553 SYPPEIKIEGLLYKKNPIYESSVMYTLDEBNTAKETTKOLNPTGKFKVSHLYDV--- 609
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
0Y 610 IKLNAAKNMILIRDRKFHYDRNNIIVAGADESVYKAEHREYVINSSTEG-----LL 657
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 610 -KLTPKKNVYTIK--LSIYDND---AESNDNSIGKWTNTNIVSGGNGKQYSSNPNANLT 664
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
0Y 658 LNID-----KDIRKIISGYVIELEDEBGLKEYINDYEDMLNIS--SLRQGGKFFIDEXKY 710
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 665 LNTDQOEKLNKKRDIYISLWIKSEKNTQOCEITTDGETIYPTTYKTVANNKKNYKRLDIAH 724
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
0Y 711 NDKLPLYISNP 721
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 725 NIK-----SNP 730
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

RESULT 13
US-08-470-566B-5
: Sequence 5, Application us/08470566B
: Patent No. 5872212
:
GENERAL INFORMATION:
: APPLICANT: Warren, Gregory W
: APPLICANT: Koziel, Michael G
: APPLICANT: Mullins, Martha A
: APPLICANT: Nye, Gordon J
: APPLICANT: Carr, Brian
: APPLICANT: Desai, Nalin M
: APPLICANT: Kostichka, N. Kristy
: APPLICANT: Duck, Nicholas B
: TITLE OF INVENTION: No. 5872212zel Pesticidal Proteins and Strains
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
:
ADDRESSEE: No. 5872212artis Corporation
: STREET: 3054 Cornwalis Road
:

```

```

1 CITY: Research Triangle Park
2 STATE: NC
3 COUNTRY: USA
4 ZIP: 27709
5
6 COMPUTER READABLE FORM:
7 MEDIUM TYPE: floppy disk
8 COMPUTER: IBM PC compatible
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10 SOFTWARE: Patentln Release #1.0, Version #1.30B
11
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/08/470.566B
14 FILING DATE: 06-JUN-1995
15 CLASSIFICATION: 530
16
17 PRIOR APPLICATION DATA:
18 APPLICATION NUMBER: US 08/463,483
19 FILING DATE: 05-JUN-1995
20
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: US 08/314,594
23 FILING DATE: 09-SEP-1994
24
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: US 08/218,018
27 FILING DATE: 23-MAR-1994
28
29 PRIOR APPLICATION DATA:
30 APPLICATION NUMBER: US 08/037,057
31 FILING DATE: 25-MAR-1993
32
33 ATTORNEY/AGENT INFORMATION:
34 NAME: Meigs, J. Timothy
35 REGISTRATION NUMBER: 38,241
36
37 REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4
38
39 TELECOMMUNICATION INFORMATION:
40 TELEPHONE: 919-541-8587
41 TELEFAX: 919-541-8689
42
43 INFORMATION FOR SEQ ID NO: 5:
44 SEQUENCE CHARACTERISTICS:
45 LENGTH: 884 amino acids
46 TYPE: amino acid
47 TOPOLOGY: linear
48
49 MOLECULE TYPE: protein
50
51 US-08-470-566B-5
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
9
```


Db 725 NIK-----SNP 730

RESULT 15
US-09-300-529-5
Sequence 5, Application US/09300529
Patent No. 6066783
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Koziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: Genes Encoding Insecticidal Proteins
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6066783artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/300, 529
FILING DATE: TBA
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/469,334
FILING DATE: 06-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-19506L
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-300-529-5

Query Match 20.8%; Score 783.5; DB 3; Length 884;
Best Local Similarity 29.6%; Pred. No. 4,2e-50;
Matches 234; Conservative 133; Mismatches 293; Indels 131; Gaps 29;

QY 1 MKRKVLIPILMALSTILVSS--TGNLEVIQAEVK-----OENRLNSESSESSOGLG 51
DB 1 MKMKKLASVYTCITLAPMFLANGVNAVYADSKTQISTOKNO---OKENDRKGLG 57

QY 52 YPSDLNFOAPMVVTSSTGDLSPSSLELN--IPSENQYFOASINSGFIKVKSDPYEFA 109
DB 58 YFKGKDF-SNLTMFAPTRSTLIYDQGTANKKLDKKQGEYQSIWIGLQSKSETGDFFN 116
QY 110 TSADNHVIMVDDQEVINKASNSNKTREKGRLYQIKIOYQRENPTEKGLD-----FKL 163
DB 117 LSEDEQALIEINGKIISNKGKQOVVHLEKGLVPIKIEYQSD--TKFNIDSKTEKEL 174
QY 164 YWTDSONKKEVISDNLOLPELKOKSS-----NSRKRSTSGPTVDPDROND 210
DB 175 FKIDSONPOVOOQDELNPDEPNKKEQEFLLAKPSKINLFTQMKREIDED--TDTQSD 231
QY 211 GIPDSLEVEGYVDVKNKRTFLSPWISNIHERKGLTKYKSEKSTASDPYSEKYG 270
DB 232 SIPDLMEENGTYL----QNRIVKMDSL-ASKGYTKFVSNPLESTVGDPYTDYEKAR 286
QY 271 RIDKNVSPARHPVLAAYIVVDMENITLSKNEODSTQNTDSETRTTSKNTSTSRHTS 330
DB 287 DDLISNAKETFPVLAAPPSVNVSMKVLSPNENLS----- 323
QY 331 EVHGADEVHANTSTSRHTSEVHGADEVH-----AVAIHSLAGERTMETMG-- 380
DB 324 ---NSVESHSTNMSTNTTEGASVEAGICPKISRGVSVNYOHSYAOE--WGTSTGNT 378
QY 381 --LNTADTARLANIRVYNTGTAPIYVNLPTTSLVLGKQNTLATIKAKENQLSQILAPNN 438
DB 379 SOFNASAGYLANVYNNVNGTAYDKPTSPVL--NNDTITATITAKSNSPALNISPE 437
QY 439 YPSKRLAIALMADDDSEPTITMNYNOLFLEKTKQRLDTPDQYGIATYNNENGRV 498
DB 438 STPKGQNGCIATISMDENSHPTLNKKQVDNLNNKPMLETNOTDG--YKIKIDHG 494
QY 499 RVDGSMNSEVLPOIETTARIIFNCKDLNVERIAAVNPDPLETPKPDMTLKEALKT 558
DB 495 NIVTGEHNGVYQOIKAKTASIIYDDGE--RVAEKRYAANDYENPEDKT--PSLTLDALK 552
QY 559 AF--GFNEPENGILQYOGKDIETEDF--NPDQOTSONIKNOALAE-----NATNIYVLDK 609
DB 553 SVPEIKETIEGLLYKKNRPYESSVMTYLDENTAKEVQTQNDTQGFQVSHLYDV--- 609
QY 610 IKLNKMNILIDDKRHHYRNNAIAGADSVYKKAHREYNSTGDS-----LL 657
DB 610 -KLTPKMNVTIK-LSTLYDN--AESNDNSIGKWTNTIVSGNNGKKOYSSNNDPANT 664
QY 658 LNIID-----KDIRKILSGYIEIEDTEGLKEVINDRYMLNIS--SLRODGKTFIDPKKY 710
DB 665 LNTDQAEKLNKRDYIISLYMKSEKNTQCEIYIDSEIYITTKTYVNVKNDYKRIIDIAH 724
QY 711 NDKLPLYISNP 721
DB 725 NIK-----SNP 730

Search completed: December 2, 2001, 13:49:43
Job time: 223 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:48:31 : Search time 112.1 Seconds
(without alignments)
485.672 Million cell updates/sec

Title: US-09-747-521-4_COPY_1_735

Perfect score: 3765
Sequence: 1 MKRRVLPALMALSTILVSS.....LYISNPKNVNYATKENT 735

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_1101: *
1: /SIDS8/gcgdata/geneseq/AA1980.DAT: *
2: /SIDS8/gcgdata/geneseq/AA1981.DAT: *
3: /SIDS8/gcgdata/geneseq/AA1982.DAT: *
4: /SIDS8/gcgdata/geneseq/AA1983.DAT: *
5: /SIDS8/gcgdata/geneseq/AA1984.DAT: *
6: /SIDS8/gcgdata/geneseq/AA1985.DAT: *
7: /SIDS8/gcgdata/geneseq/AA1986.DAT: *
8: /SIDS8/gcgdata/geneseq/AA1987.DAT: *
9: /SIDS8/gcgdata/geneseq/AA1988.DAT: *
10: /SIDS8/gcgdata/geneseq/AA1989.DAT: *
11: /SIDS8/gcgdata/geneseq/AA1990.DAT: *
12: /SIDS8/gcgdata/geneseq/AA1991.DAT: *
13: /SIDS8/gcgdata/geneseq/AA1992.DAT: *
14: /SIDS8/gcgdata/geneseq/AA1993.DAT: *
15: /SIDS8/gcgdata/geneseq/AA1994.DAT: *
16: /SIDS8/gcgdata/geneseq/AA1995.DAT: *
17: /SIDS8/gcgdata/geneseq/AA1996.DAT: *
18: /SIDS8/gcgdata/geneseq/AA1997.DAT: *
19: /SIDS8/gcgdata/geneseq/AA1998.DAT: *
20: /SIDS8/gcgdata/geneseq/AA1999.DAT: *
21: /SIDS8/gcgdata/geneseq/AA2000.DAT: *
22: /SIDS8/gcgdata/geneseq/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3652	97.0	764	AA56958	B. anthracis prote
2	3652	97.0	764	AA56958	Wild type B. anthr
3	3551.5	94.3	763	AA56960	B. anthracis TPA-P
4	3519	93.5	735	AA60179	Protective antigen
5	3519	93.5	736	AA56959	B. anthracis MAT-P
6	3519	93.5	903	AA60183	PA(1-725)-----Huma
7	3342	88.8	719	AA60193	Modified protectiv
8	2664	70.8	569	AA56961	B. anthracis PA63
9	787.5	20.9	884	AA91239	B. cereus VIP1A(a)
10	783.5	20.8	884	AA19509	B. cereus VIP1A(a)
11	783.5	20.8	884	AA46712	100 kDa VIP1A(a) t

12	783.5	20.8	1346	AA91245	VIP2A(a) and VIP1A
13	783.5	20.8	1346	AA19513	B. cereus VIP1A(a)
14	783.5	20.8	1346	AA46723	VIP1A(a)/VIP2A(a)
15	781.5	20.8	880	AA60224	Bacillus thuringie
16	781.5	20.7	881	AA59277	MIS toxin from B.
17	778.5	20.7	884	AA63793	Bacillus cereus 10
18	776.5	20.6	784	AA82948	MIS toxin of Bacil
19	771	20.5	1338	AA91247	VIP2A(a)-VIP1A(a)
20	771	20.5	1338	AA19520	Maize optimised-B.
21	771	20.5	1338	AA46731	VIP2A(a)/VIP1A(a)
22	769	20.4	852	AA91246	VIP1A(a) protein w
23	769	20.4	852	AA19516	Maize optimised-B.
24	769	20.4	852	AA46727	Maize optimised VI
25	766.5	20.4	860	AA59282	MIS-8 toxin from B
26	755.5	20.1	834	AA91242	B. thuringiensis V
27	755.5	20.1	834	AA19512	B. thuringiensis V
28	755.5	20.1	834	AA46722	Vegetative insect
29	596.5	15.8	425	AA60222	Bacillus thuringie
30	584	15.5	667	AA91240	B. cereus VIP1 pro
31	580	15.4	667	AA19510	B. cereus 80 kD VI
32	580	15.4	667	AA46713	80 kDa VIP1A(a) to
33	578	15.4	667	AA63794	Bacillus cereus 80
34	541	14.4	357	AA60228	Bacillus thuringie
35	538	14.3	357	AA60231	Bacillus thuringie
36	521	13.8	357	AA60218	Bacillus thuringie
37	521	13.8	357	AA60226	Bacillus thuringie
38	503	13.0	357	AA60227	Bacillus thuringie
39	450	12.0	327	AA60220	Bacillus thuringie
40	447	11.9	327	AA60223	Bacillus thuringie
41	427.5	11.4	348	AA60219	Bacillus thuringie
42	425.5	11.3	348	AA60229	Bacillus thuringie
43	419.5	11.1	347	AA60221	Bacillus thuringie
44	419.5	11.1	347	AA59276	Toxin from B. thur
45	396	10.5	333	AA60230	Bacillus thuringie

ALIGNMENTS

RESULT	ID	AA56958	standard; protein; 764 AA.
XX	AA56958:		
AC	25-APR-2000	(first entry)	
DT			
XX			
DE			B. anthracis protective antigen (PA) protein.
XX			
KW			Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA;
KW			tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.
XX			
OS			Bacillus anthracis.
XX			
PN	WO200002522-A2.		
XX	20-JAN-2000.		
PD			
XX			
PF	09-JUL-1999:	99WO-US15568.	
XX			
PR	10-JUL-1998:	98US-0092416.	
XX			
PA	(USME-) US MEDICAL RES INST INFECTIOUS DISEASES.		
XX			
XX			
PI	Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;		
XX			
DR	WPI: 2000-182165/16.		
XX	N-PSDB: AA256874.		
XX			
PT	Recombinant DNA construct useful as vaccines for anthrax, in producing		
PT	host cells for analyzing the drugs and agents inhibiting anthrax		
XX			
PS	Disclosure: Page 33: 35pp: English.		

XX The invention provides a recombinant DNA construct that comprises a
 CC vector and at least one nucleic acid (or its fragment) encoding a
 CC combination of Bacillus anthracis proteins, selected from protective
 CC antigen (PA), Mat-PA (PA with its secretory signals removed), rPA-PA
 CC (PA with its secretory signals replaced with those of tissue plasminogen
 CC activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine
 CC for anthrax and in producing infectious alpha virus particles. These
 CC particles, expressing the B. anthracis proteins are useful also as
 CC vaccines for anthrax. Host cells transformed with the construct are
 CC useful for analyzing the effectiveness of drugs and agents that inhibit
 CC anthrax or B. anthracis proteins. The present sequence represents a
 CC B. anthracis PA protein.

XX Sequence 764 AA:

Query Match 97.0%; Score 3652; DB 21; Length 764;

Best Local Similarity 97.6%; Pred. No. 5.9e-237; Mismatches 16; Indels 0; Gaps 0;

Matches 717; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

1 MKRRVLPIMLSTIVSSNGLEVIQAEVKQENRLNESSSSQGLGYFSDLNFOA 60
 1 mkkrrvlpimlslstlvssngleviqaevkqenrlnesesssqgllgyfstdlnfqa 60

61 PMVVTSSSTGDISPSSELENIPSENOYFQSAIWSGFVKKSDEYTFATSADNHVTMV 120
 61 pmvvtssstgdispsselelenipseenyfqsaiswgfivkksdeytfatsadnhvtmv 120

121 DDOEVYNKASNNKIRLEKGRLYQIKYQRENPTREKGFELWYSDSNKKEVSSDL 180
 121 ddoevynkasnnkirllekgrlyqikyqrenptrekgfelfwysdnkkekessdl 180

121 ddoevynkasnnkirllekgrlyqikyqrenptrekgfelfwysdnkkekessdl 180

181 QLPBLKOKSSNRKRSTSGPTVDRDNDGIPDSLEVGYYVDVKNKRTFSPMISNH 240
 181 qlpblkoksnsnrkrstsgptvdrdndgipdslevgyyvdvknkrtfspmish 240

181 qlpblkoksnsnrkrstsgptvdrdndgipdslevgyyvdvknkrtfspmish 240

241 EKKGLTKTKSSPEKKSTASDPYSDEKVTGRIDKNVSPARHPLVAAYPIVHDMENIL 300
 241 ekkgltktksspekstasdpysdekvtgridknvsparhplvaaypivhdmenil 300

241 ekkgltktksspekstasdpysdekvtgridknvsparhplvaaypivhdmenil 300

301 SKNEOSQNTDSEMTRTISKNTSTSRHTSEVHNAEVAHNTSTSRHTSEVHNAEVA 360
 301 skneosqntdsemtstskntstsrhtsevhnaevahntstsrhtsevhnaevah 360

301 skneosqntdsemtstskntstsrhtsevhnaevahntstsrhtsevhnaevah 360

361 VAIDHSLSIAGERTWAETGLTADTARLANIRYVNGTAPYVNLPTTSLVLGKQTL 420
 361 vaidhslsiagertwaetgladtarlaniiryvngtapyvnlpttsslvgkqtl 420

361 vaidhslsiagertwaetgladtarlaniiryvngtapyvnlpttsslvgkqtl 420

421 ATIRAKENOLSIAPNNYPSKNIAPIALNAODESSPTTMTNTOELEKTKQLRD 480
 421 atirakenolsiapnnypskniapiaalnaodesstptmtntoelkqtlrd 480

421 atirakenolsiapnnypskniapiaalnaodesstptmtntoelkqtlrd 480

481 TDQVGNATYFENGRRVVDFTGSMNSEVLPOIETTAIINGDNLNVEERIAAVNS 540
 481 tdqvgnatyfenrrvvdftgsmnssevlpoiettaiingdnlnevriaavns 540

481 tdqvgnatyfenrrvvdftgsmnssevlpoiettaiingdnlnevriaavns 540

541 DPLETTKPDMLTEALKAIFGNEPNGNLQYQKDIETEFDFNFDOOTSNIKQLAELNA 600
 541 dplettkpdmeltealkaifgnepngnlqyqkdietefdfnfnfdootsnikqlaelna 600

541 dplettkpdmeltealkaifgnepngnlqyqkdietefdfnfnfdootsnikqlaelna 600

601 TNYIVVLDKIKLNAMKNIIRDKRPHYDRNNIAGADESVKKAHREVIINSSTEGILLNI 660
 601 tnyivvldkiklnamkniiirdkrphydrrnniagadesvkkahreviinsstegillni 660

601 tnyivvldkiklnamkniiirdkrphydrrnniagadesvkkahreviinsstegillni 660

661 DKDIRKIIISGYIYEIETPGEKKEVINDRYDMNISLRDGGTFIDFKYNNKPLYSIN 720
 661 dkdirkiiisgyiyeietpgekkevindrydmnislrdggtfidfkynnkplysin 720

661 dkdirkiiisgyiyeietpgekkevindrydmnislrdggtfidfkynnkplysin 720

721 PNYKVNVYAVTKENT 735
 721 pnykvnyvavtkent 735

721 pnykvnyvavtkent 735

RESULT 2

AB47306
 ID AB47306 standard; Protein; 764 AA.

XX AB47306;

XX 29-AUG-2001 (first entry)

XX Wild type B. anthracis protective antigen.

XX Lethal factor; LF; immunogen; LF4; protective antigen; PA; DNA vaccine;
 XX humoral; cell-mediated; immune memory response.

XX Bacillus anthracis.

XX Key Location/Qualifiers

XX Peptide 1..29 /label= Signal peptide

XX Protein /note= "Not given in the specification"

XX Peptide 30..764 /label= PA

XX W0200145639-A2.

XX 28-JUN-2001.

XX 21-DEC-2000; 2000WO-US34912.

XX 22-DEC-1999; 99US-0171459.

XX (OHIS) UNIV OHIO STATE RES FOUND.

XX (GALLI/) GALLOWAY D R.

XX (MATEC/) MATECZUN A J.

XX Galloway DR, Mateczun AJ;

XX WPI; 2001-408540/43.

XX N-PSDB; AAC86016.

XX Claim 5; Fig 2; 33pp; English.

XX This sequence shows the B. anthracis protective antigen (PA). An
 CC immunogenic fragment of PA, pCPA, can be used to produce an immune
 CC response which protects an animal against lethal infection with
 CC Bacillus anthracis. DNA encoding the B. anthracis PA can be used
 CC in conjunction with DNA encoding the lethal factor (LF) in a
 CC DNA vaccine. Using a DNA vaccine which encodes the mutated LF protein
 CC or fragment alone or in combination with a DNA encoding the PA protein
 CC or its fragment, both components (humoral and cell-mediated) of the
 CC immune system are stimulated, which results in longer term immune
 CC memory response. The combined use of a mutated LF and PA gene or their
 CC fragments results in a higher level of immune response, as judged by
 CC overall serum antibody titers for LF and PA antigens, than the use of
 CC either LF or PA genes in separate immunizations.

XX Sequence 764 AA:

Query Match 97.0%; Score 3652; DB 22; Length 764;

Best Local Similarity 97.6%; Pred. No. 5.9e-237; Mismatches 16; Indels 0; Gaps 0;

Matches 717; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

1 MKRRVLPIMLSTIVSSNGLEVIQAEVKQENRLNESSSSQGLGYFSDLNFOA 60
 1 mkkrrvlpimlslstlvssngleviqaevkqenrlnesesssqgllgyfstdlnfqa 60

Db 602 lylvtdkklhnmlllrdrfhydrmlavagadesvakeahrevlnssteglllnldk 661
QY 663 DIRKILSGYIIEDETEGLKEVINRDMNLNISRQDGKPTIDFKKYNDKRLPLYISPN 722
Db 662 dirklslsgylveiedteglkevindrmdlnisslrqdgkftldkkyndkrlplyisnpn 721
QY 723 YKVNVAVTKEMT 735
Db 722 ykvnvyavtkent 734

RESULT 4

AAR60179 standard; protein: 735 AA.
ID AAR60179 standard; protein: 735 AA.

AAR60179:

03-APR-1995 (first entry)

Protective antigen of Bacillus anthracis.

Anthrax: Bacillus anthracis; fusion protein; protective antigen;

protective antigen; cell killing; targeting; targeting; pathogen;

intracellular; HIV; human immunodeficiency virus; toxin.

Bacillus anthracis.

MO9418332-A.

14-FEB-1994: 94WO-US01624.

12-FEB-1993: 93US-0021601.

25-JUN-1993: 93US-0082849.

(USSH) US DEPT HEALTH & HUMAN SERVICES.

Atora N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;

WPI: 1994-279753/34.

N-PSDB: AAQ70180.

Nucleic acid encoding anthrax toxin fusion protein - useful for

targeting toxin to specific cells, eg for killing tumour cells

or HIV-infected cells

or HIV-infected cells

or HIV-infected cells

or HIV-infected cells

or HIV-infected cells

or HIV-infected cells

or HIV-infected cells

or HIV-infected cells

or HIV-infected cells

or HIV-infected cells

or HIV-infected cells

or HIV-infected cells

or HIV-infected cells

or HIV-infected cells

QY 150 QRENTEKGLDFKLYWTSQNKKEVISSDNLQPELKOKSSNRKRKSTAGPTVPDRDN 209
Db 121 gremptekyldfklywtsqpkkevissdnlqpepkksnsrkrkstagsptvdrdn 180
QY 210 DGIPSLVEEGTVVYKRRFTLSFWISNIEKKGLTRYKSSPEKWSASDPYSDFEYV 269
Db 181 dgipslveegtyvdvkkrrftlspwshnhekgltryksspekwsasdpysdfekvt 240
QY 270 GRIDKNVSEARHPPLVAAYPIVHVMENIILSKNDSOSTONTSEPTISKNTSRPT 329
Db 241 gridknvsearhplvaaypivhvmeniilsknndstontseptiskntsrpt 300
QY 330 SEVHGNAEVHANTSTSRPTSEVHGNAEVAVIDHSLSLAGEPTAETMGLNTADTARL 389
Db 301 sevhgnaevhasffdiggsvasgfenssvaidhsislagertwaetmglntadtarl 360
QY 390 NANIRYVNTGTAPRYNVLPYTSVLGKKNQTLATIRAKENQSLIAPNNYPPSKLARIA 449
Db 361 nanirylvntgtaprynvlpysvlgkknqtlatlakenglsqllapnnyppsklaria 420
QY 450 LNAQDFSSFTITMNYNOFLELEKTQRLDPTDQVYGNATYNEENGVRVDTGSMSEV 509
Db 421 lnaqdfsstptitmyngflekckqlriddqygniatyngfengrvvdtgsnsew 480
QY 510 LPQIETARILFNGKDLNVERRIAAYNPSPLETTRKPMTLKALKIAGFNEPNSNL 569
Db 481 lpqietarilifngkdlnverriaavnpdsplettrkpmtlkealkiagfnepnsl 540
QY 570 QYOGKITEFPNPOORSONIKNOELANPTNYTVLDTIKLANKMILIRDRRFHYDR 629
Db 541 qyogkitlefndtqgtsqknglaelnatnlylvtdkklhnmlllrdrfhydr 600
QY 630 NNIAVGADESYYKEARREVINSSTEGLLNIDDKIRKILSGYIIEDETEGLKEVINR 689
Db 601 nniavgadesvakeahrevinssteglllnidkdrklslsgylveiedteglkevindr 660
QY 690 DMLNISRQDGKPTIDFKKYNDKRLPLYISNPKNVVAVTKEMT 735
Db 661 dmlnisslrqdgkftldkkyndkrlplyisnpknvyavtkent 706

RESULT 5

AAY56959 standard; protein: 736 AA.
ID AAY56959 standard; protein: 736 AA.

AAY56959:

25-APR-2000 (first entry)

B. anthracis MAT-PA protein.

Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA;

toxic plasmidogen activator; PA63; vaccine; anthrax; antibacterial.

Bacillus anthracis.

WO200002522-A2.

20-JAN-2000.

09-JUL-1999: 99WO-US15568.

10-JUL-1998: 98US-0092416.

(USME-) US MEDICAL RES INST INFECTIOUS DISEASES.

Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;

WPI: 2000-182165/16.

N-PSDB: AAZ56875.
Recombinant DNA construct useful as vaccines for anthrax, in producing

host cells for analyzing the drugs and agents inhibiting anthrax

PT

XX Disclosure; Page 34; 35pp; English.

PS The invention provides a recombinant DNA construct that comprises a
XX vector and at least one nucleic acid (or its fragment) encoding a
CC combination of Bacillus anthracis proteins, selected from protective
CC antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA
CC (PA with its secretory signals replaced with those of tissue plasminogen
CC activator) and PA63 (63 KDa PA). The construct is useful as a DNA vaccine
CC for anthrax and in producing infectious alpha virus particles. These
CC particles, expressing the B. anthracis proteins are useful also as
CC vaccines for anthrax. Host cells transformed with the construct are
CC useful for analyzing the effectiveness of drugs and agents that inhibit
CC anthrax or B. anthracis proteins. The present sequence represents a
CC B. anthracis MAT-PA protein.

XX Sequence 736 AA;

Query Match 93.5%; Score 3519; DB 21; Length 736;
Best Local Similarity 97.5%; Pred. No. 4,8e-228;
Matches 688; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 30 EYKQENRLNESESSQGLGYFSDNFQAPWVVSSTTGDISPSELENIPSENOYF 89
DB 2 evkqenrlneesssgqllgyfsgdnfqapwvvssttgdispselenipseqnyf 61
QY 90 QSAIWSGFIKVKKSDEYTFATSADNHVTMWVDDQEVINKASNSNKIRLEKGRLYQIKIY 149
DB 62 qsalwsgfikvkksdeytfatsadnhvtmwddqevinkasnsnkirleqgrlyqikly 121
QY 150 QRENPFKGLDFKLYWTDSONKKEVSSDNLQPELKOKSSNRKRSTAGPTVPDRN 209
DB 122 qrenpfekgldfklywtdsqnkkvssdnqlpelkqksnsrkrstagsptvpdrn 181
QY 210 DGIPDSLEVEGYVDVKNKRTFLSPWISNHEKKGLTKKSSPEKMTASDPYSDEKYT 269
DB 182 dgipdslevegylvdvknkrftflspwismhekkgltkksspekstaadpysdekvt 241
QY 270 GRIDKNVSPRARPLVAAPYIVHVDMENIILSKNEDOSTONTSEFTISKNTSTSRTH 329
DB 242 gridknvspearplvaapyivhvdmenilsknedqstqntdsetrlskntstsrth 301
QY 330 SEVHGAEEVHANTSTSRTHSEVHGAEEVHVAIDHSLAGERTWAEWTGLTADTARL 389
DB 302 sevhgaeevhasftdigsvsagfnsnsstvaldhsislaager twaelmgltadtarl 361
QY 390 NANIRVTNTGTAPYIVNVLPFTSLVGNQTLATIKAKENQLSQILAPNNYPSKNLAPIA 449
DB 362 nanirvntgtaplynvlpftslvgnqtlatikakenglsqilapnnypsknlapia 421
QY 450 LNAQDFFSSTPTTMVNOFLELEKTRQRLDQVYGNATYVFNENGRVAVDGTGNMSEV 509
DB 422 lnaqddfsstpttmvnoflelektqrltdqvygnatynfengrvavdgtgnmsev 481
QY 510 LQIQIOTTRIIIFNGKDLNVERRIAANVPSDPLETTKTPMTLKEALKIAFGNENRGNL 569
DB 482 lpiqiottriiifngkdlnverriaavnsdpdlectktpmtlkealkiafgnenrgnl 541
QY 570 QYQKIDITEFDNFDOOTSQNIKNQLAELNATNYIVLQIKILAKAMNIIIRKRFHYDR 629
DB 542 qyqkditfednfdootsqniknqlaelnatiyivldikilakamniiirkrfhydr 601
QY 630 NNIANGADESVKAEHREVINSSTEGLLNIDKIRKILSGYVEIETEGLEKLEVIANDRY 669
DB 602 nniavgadesvvkeahrevinsstegllnidkirkilsgyveietegglekleviandry 661
QY 690 DMNISLSRODGKTFIDFKKYNKLPYISNPKYKVNVAVTKEMT 735
DB 662 dmnisslrtdgkftidfkknkldpyisnpkykvnvavtkemt 707

RESULT 6

AAR60183
ID AAR60183 standard; Protein; 903 AA.

XX AAR60183;
AC 04-APR-1995 (first entry)
DT PA(1-725)-----Human CD4 fusion protein coding sequence.

XX Anthrax; Bacillus anthracis; fusion protein; lethal factor;
KW protective antigen; cell killing; targeting; targeting; pathogen;
KW intracellular; HIV; human immunodeficiency virus; toxin;
KW pseudomonas; exotoxin.

XX Bacillus anthracis.
OS Homo sapiens.

XX W09418332-A.

XX 18-AUG-1994.

XX 14-FEB-1994; 94WO-US01624.

XX 12-FEB-1993; 93US-0021601.

XX 25-JUN-1993; 93US-0082849.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Atora N, Kimpel K, Leppla SH, Nichols PJ, Singh Y;

XX WPI; 1994-279753/34.

XX N-PSDB; AAO70184.

XX Nucleic acid encoding anthrax toxin fusion protein - useful for
PT targeting toxin to specific cells, eg for killing tumor cells
PT or HIV-infected cells

PS Disclosure; Page 100-103; 124pp; English.

XX This sequence is a fusion protein comprising amino acid residues
CC 1-725 of the anthrax protective antigen protein and residues 1-178
CC of human CD4, the portion which binds to gp120 on HIV infected
CC cells. Such fusion proteins may be useful for the specific killing
CC of tumor cells or the killing of cells infected with intracellular
CC pathogens, especially HIV, depending on their components.

XX Sequence 903 AA;

Query Match 93.5%; Score 3519; DB 15; Length 903;
Best Local Similarity 97.5%; Pred. No. 6,5e-228;
Matches 688; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 30 EYKQENRLNESESSQGLGYFSDNFQAPWVVSSTTGDISPSELENIPSENOYF 89
DB 1 evkqenrlneesssgqllgyfsgdnfqapwvvssttgdispselenipseqnyf 60
QY 90 QSAIWSGFIKVKKSDEYTFATSADNHVTMWVDDQEVINKASNSNKIRLEKGRLYQIKIY 149
DB 61 qsalwsgfikvkksdeytfatsadnhvtmwddqevinkasnsnkirleqgrlyqikly 120
QY 150 QRENPFKGLDFKLYWTDSONKKEVSSDNLQPELKOKSSNRKRSTAGPTVPDRN 209
DB 121 qrenpfekgldfklywtdsqnkkvssdnqlpelkqksnsrkrstagsptvpdrn 180
QY 210 DGIPDSLEVEGYVDVKNKRTFLSPWISNHEKKGLTKKSSPEKMTASDPYSDEKYT 269
DB 181 dgipdslevegylvdvknkrftflspwismhekkgltkksspekstaadpysdekvt 240
QY 270 GRIDKNVSPRARPLVAAPYIVHVDMENIILSKNEDOSTONTSEFTISKNTSTSRTH 329
DB 241 gridknvspearplvaapyivhvdmenilsknedqstqntdsetrlskntstsrth 300

Qy	330	SEVGNNAEVAHANTSTSRHHSSEVGNABEVAHVAIDHSLSLAGERTMAETMGLTADTARL	369
Db	301	sevgnaevhasifdldgsvsaqfismnsclvaidsbslslagertwetmgltadtlarl	360
Qy	390	NANRYVNTGAPILYNNLPPTSLVLGNKOTLARIKAKENOLSOLIAPNNTYPPSKNLAPIA	449
Db	361	nanlryntgapiynvlpptslylgnkqtlatlkaenqsgllapmnypsknlapia	420
Qy	450	LMAODDESSPTITMNYNOFLELEKTKOLRLDPTDQVYGNIAITYNEENGVRVDTGSMNSEV	509
Db	421	lnaqqdfessptiltnmyqfielektqrltdtdqvygniatyngfengrvvdctgsnwsew	480
Qy	510	LPOIETPARILIFNGKOLNLYERRILAAVNPSPDETTRKPMWLTKEALKAFGRNEPENGUL	569
Db	481	lpqldetarlilfngkdlnlverriaavmpsdplettkpmllkealkiafgfnepgnl	540
Qy	570	QYOGKDIEFPNFDQOSTONIKNOLELANTNTYLYDLKILKAKKNNLILROKRFHYDR	629
Db	541	qygqkditeffnfdqgtsqnlkqlelnaetnlyvtdklklnakmllllrdkrfhydr	600
Qy	630	NNIAVGADESVKAEHREVINSTEGGLNLIDKDIRKISGYIYEIEDTEGLKEVINDRY	689
Db	601	nniavgadevsvkaehrevlnsstegllllddkdirklsgylveledteglkevindry	660
Qy	650	DLMLNSSLRODGKTFIDFKKXNDKLPILYSINPKXNVAVYATKENT	735
Db	661	dmlnsslrgdgkffidfkkyndklpilyisnpmkxnvvyavtkent	706

```

RESULT 7
ID AAR60193
XX AAR60193 standard; Protein: 719 AA.
XX
XX AAR60193:
XX
XX 04-APR-1995 (first entry)
XX
XX Modified protective antigen of Bacillus anthracis.
XX
XX Anthrax; Bacillus anthracis; fusion protein; lethal factor;
XX protective antigen; cell killing; targeting; pathogen;
XX Intracellular; HIV; human immunodeficiency virus; toxin.
XX
XX Bacillus anthracis.
XX
XX WO9418332-A.
XX
XX 18-AUG-1994.
XX
XX 14-FEB-1994; 94WO-US01624.
XX
XX 12-FEB-1993; 93US-0021601.
XX
XX 25-JUN-1993; 93US-0082849.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Atora N, Klimpel K, Leppla SH, Nichols PJ, Singh Y:
XX
XX WPI: 1994-279753/34.
XX
XX N-PSDB; AAO70189.
XX
XX Nucleic acid encoding anthrax toxin fusion protein - useful for
XX targeting toxin to specific cells, eg for killing tumour cells
XX or HIV-infected cells
XX
XX Example 6; Page 114-115; 124pp; English.
XX
XX The sequence encoding the protective antigen of Bacillus anthracis
XX may be used in the construction of a nucleic acid which encodes a
XX fusion protein comprising the anthrax protective antigen binding
XX domain of the native anthrax lethal factor and a sequence encoding an
XX activity inducing domain of a second protein. Such fusion proteins
XX are useful for the specific killing of tumour cells or the killing of
XX

```

CC cells infected with intracellular pathogens, especially HIV,
CC depending on the second component. The protective antigen and other
CC toxins require proteolytic cleavage to acquire activity. Since some
CC cells infected with an intracellular pathogen possess an active
CC protease with quite a narrow substrate specificity e.g. HIV, the
CC protease cleavage site found in the native toxin is replaced with an
CC intracellular pathogen specific protease site (See AAR60184-88). The
CC protease in cells that are infected with an intracellular pathogen
CC cleaves the modified toxin which is then rendered active and kills
CC the cell. This sequence is a modified Bacillus anthracis protective
CC antigen which has the amino acids originally at positions 162-171
CC replaced with the HIV protease cleavable sequence described in
CC AAR60186.
XX

Sequence 719 AA:
SQ

Query Match	88.8%	Score 3342	DB 15	Length 719
Best Local Similarity	92.8%	Pred. No. 3.6e-216		
Matches 659	Conservative 6	Mismatches 21	Indels 24	Gaps 2
QY	30	EVKQNRRLNESSESSOGLGYGFSDLNFQAPMVVTSSTPGDLSIPSSLENIIPSENOYF	89	
Db	1	evkqgnrlinesesssgllgyfisdlnfqapmvvtstgtglslpsselenipengyf	60	
QY	90	QSAIWSGFIKKVKSDEPTFATSDAHNYVMVDDQEVINKAKSNKIRLEKGLYQIKIOY	149	
Db	61	qsaiwsqfkkvksdeyftatadnhvmvddqevinkasnskirklekgllyqikioy	120	
QY	150	QRENPTKGLDEPKLWMTPOSQNKKEVSISSDNLOLPELKQSNSS---RKRSSTASGPRV	205	
Db	121	qrenptekgldekllwmtposqokkevslsdnqlprlkqksntatelmrgnflagprcvr	180	
QY	206	DRDNGCITDLSLEVEGYTDVKKKRFPLSPWISINHEKKGLTYKSSPEKMSASDPYSDF	265	
Db	181	drdngitdpsleveytdvknkrftflspwlsnlhekkxgltkykspskwsasdpysdf	240	
QY	266	EKVTRIDKKNPSPELRHPVLAAYPIVHVDMENIILSKNEDOSQTQMTDSETRTISKNTSTS	325	
Db	241	ekvtgtridknvspearhpvrvaaypvlvhvdmennllsknegdqctqtdsetrtlskntsts	300	
QY	326	PRHTSEVHGNMAEVHANTSTSRHTSEVHGNMAEVHAYVDHDSISLAGEPRMATMGLNTAD	385	
Db	301	rhtsevhgnmaevhantstsrhtsevhgnmaevhayaivdhdsislageprmatmglntad	360	
QY	386	TARLANIRYVVTGAPRIYVNPPTTSILYLGKKQOTLATIKAKENQLSQILAPANNYPSPKNL	445	
Db	361	tarlanirlyvvtgarpriyvnppttsilylgkqatlatikakenqslsqilapannypsknl	420	
QY	446	APIALNADDESSPTITMNYNOFLELEKTKOLRLDQOYGYAIATYNEBNGVRVDTGSN	505	
Db	421	apialnagddfsptitlmn-----ygnlaeylnfengrvrvtgtsn	460	
QY	506	WSEVHPQIOETARIIIFNGKDLNLVERRIAAVNPSPDLPTTRPMDTLKALKIARGFNPR	565	
Db	461	wsevhprqioetariilfngkdlnlverriaavnpdppttrpmdtlkalkiargfnpr	520	
QY	566	NGNLOYOGKDIPEDFNFDQOTSQNIKNQDLAEIANTNTYTVLVDKIKLNAKNMILIRDKRF	625	
Db	521	ngnluygqkditefdnfedqfsgnllknqjaelnatnlytvldkiklnaknmillrdkrf	580	
QY	626	HYDRNNIIVAGADESVYKKEAHREVINSSPTGGLLNLNDKOKRLISGTYIETEGEGLKEYI	685	
Db	581	hydrnniivagadesvvykkehrevlnsstegllnlndkdkrlisgyiyeitedeglekeyi	640	
QY	686	NDRYMLNLTSSLRDQGTFTIDFKKYNNDKIRPLYTISPNKYVNVNYYAATKENT 735		
Db	641	ndrymlnltsslrqggtftidfkkyndkrlplytispnykynvnyyavtkent 690		

XX AAY56961;
 AC
 XX
 DT 25-APR-2000 (first entry)
 DE B. anthracis PA63 protein.
 XX
 KM Bacillus anthracis protein: protective antigen; PA: MAT-PA; TPA-PA;
 KM tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.
 XX
 OS Bacillus anthracis.
 XX
 PN WO200002522-A2.
 XX
 PD 20-JAN-2000.
 XX
 PF 09-JUL-1999; 99WO-US15568.
 XX
 PR 10-JUL-1998; 98US-0092416.
 XX
 PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
 XX
 PI Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;
 XX
 DR WPI: 2000-182165/16.
 DR N-PSDB; AA556877.
 XX
 PT Recombinant DNA construct useful as vaccines for anthrax, in producing
 PT host cells for analyzing the drugs and agents inhibiting anthrax -
 XX
 PS Disclosure; Page 35; 35pp; English.
 XX
 CC The invention provides a recombinant DNA construct that comprises a
 CC vector and at least one nucleic acid (or its fragment) encoding a
 CC combination of Bacillus anthracis proteins, selected from protective
 CC antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA
 CC (PA with its secretory signals replaced with those of tissue plasminogen
 CC activator) and PA63 (63 KDa PA). The construct is useful as a DNA vaccine
 CC for anthrax and in producing infectious alpha virus particles. These
 CC particles, expressing the B. anthracis proteins are useful also as
 CC vaccines for anthrax. Host cells transformed with the construct are
 CC useful for analyzing the effectiveness of drugs and agents that inhibit
 CC anthrax or B. anthracis proteins. The present sequence represents a
 CC B. anthracis PA63 protein.
 CC
 CC Sequence 569 AA;
 XX
 SQ

Query Match 70.8%; Score 2664; DB 21; Length 569;
 Best Local Similarity 96.7%; Pred. No. 8.8e-171;
 Matches 521; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 197 STSAGPTVPDRNDGIPDSLEVEGYTVVKNKRTFLSPWISNIHEKKGLTKYKSSPEKWS 256
 DB 2 stsaagptvpdrndgipdslevegtyvdknktflspwlsnhkxgltkykaspekws 61
 QY 257 TADDPYSDPEKKYVGRIDKNNSPFAHPLVAAYPIVHDMENIILSKNEDOSTQTDSETR 316
 DB 62 tasdpystdpekkvgridknnspfaahplvaaypivhdmenniilsknedostqtdsetr 121
 QY 317 TIKKNSTSTRHTSEVHGNAEVAHNTSTSRHTSEVHGNAEVAHVALDHSISLAGEETWA 376
 DB 122 tikknststrhtsevhgnaevahntstsrhtsevhgnaevahvaldhsislagetwa 181
 QY 377 ETMGLNTADTARLANIRVYNTGTAPILYVLPPTSLVLGKNQTLATIKAKENQLSQILAP 436
 DB 182 etmglntadtarlanirvnyntgtapilvlpptslvlgknqtlatkakenglsqilap 241
 QY 437 NNTYPSKNTLAPIALNADODSSPTITMNYQOLFLEKTKQLRLDDOYGINATYFENG 496
 DB 242 nntypskntlapialnadodssptitmnynqolflekktqlrldddyginatynfeng 301
 QY 497 RVRVDTSNMSEVLPQIOETTARIIFNGKDLNLVERLIAAVNPSPDLETTKPDMLKEAL 556

DB 302 rrvvdtsnmsevpqigettariifngkdnlverliaavnpdletctpdmlkeal 361
 QY 557 KIAFENPEPNCNLOYGCKDITREFDQFNSQIKNOJLAELMNTVYTLDKIKLNAM 616
 DB 362 kiafepnpgnldyqgkdltefndfngqtsqnlknlaelnatnlytvidkiklnaem 421
 QY 617 NIILROKRFHNDNRNINAVGADESIVYKKAHREVINSSTGGLNLNDKQIRKILSGYVIE 676
 DB 422 niilrdrfhydrnlavadesvveahrevlnsteglllnldkdlrkilsgylve 481
 QY 677 DTEGLKEVINDRYDMLNISSLRQDGKTFIDFKKYNDRKPLVYISNPNYVANYAATKENT 735
 DB 482 dteglkevindrmdlnisslrqdgktdfkknkdrplvysnpnyvanyavtKent 540

RESULT 9
 AAR91239
 ID AAR91239 standard; Protein; 884 AA.
 XX
 AC AAR91239;
 XX
 DT 14-AUG-1996 (first entry)
 XX
 DE B. cereus VIP1A(a) insect-specific protein.
 XX
 KW Pesticide; insecticide; biological control agent; Lepidoptera;
 KW Coleoptera; transgenic plant; maize; insect resistance;
 KW western corn rootworm; Diabrotica virgifera virgifera; VIP.
 XX
 OS Bacillus cereus strain AB78 (NRRL B-21058).
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..33
 FT /label= sig_peptide
 XX
 PN W09610083-A1.
 XX
 PD 04-APR-1996.
 XX
 XX 27-SEP-1995; 95WO-EP03826.
 PF 27-SEP-1995; 95WO-EP03826.
 XX
 PR 05-JUN-1995; 95US-0463483.
 PR 28-SEP-1994; 94US-0314594.
 XX
 PA (CIBA) CIBA GEIGY AG.
 XX
 PI Carr B, Desai NM, Duck NB, Estruch JJ, Kostichka K;
 PI Koziel MG, Mullinama, Nye GJ, Warren GW;
 XX
 DR WPI: 1996-200921/20.
 DR N-PSDB; AAT13940.
 XX
 PT Bacillus strain producing insecticidal protein during vegetative
 PT growth - used in the control of Lepidoptera and Coleoptera pests
 XX
 PS Claim 15; Page 121-124; 242pp; English.
 XX
 CC Insect-specific protein VIP1A(a) (AAR91239) of Bacillus cereus AB78
 CC shows activity against Diabrotica spp. pests such as the
 CC western corn rootworm. It is encoded by the VIP1A(a) gene (AAT13940)
 CC isolated from a cosmid clone of AB78. VIP1A(a) can be expressed in
 CC e.g. bacterial hosts to provide biological control agents having
 CC increased activity or target range, or can be expressed in transgenic
 CC plants, esp. maize, to improve insect resistance. It is preferably
 CC expressed as a fusion protein (see also AAR91245) with auxiliary protein
 CC VIP2A(a) (AAR91238).
 CC
 CC Sequence 884 AA;
 SQ

Query Match 20.9%; Score 787.5; DB 17; Length 884;
 Best Local Similarity 26.7%; Pred. No. 1.9e-44;

Matches	248;	Conservative	134;	Mismatches	291;	Indels	255;	Gaps	32;
QY	1	MKKRKVLIPLMALSTILVSS--TGNLEVIQAEVK-----OENRLNSESSESSOGILGY	51						
DB	1	mkmmkkkiasvvtctcllapmfingvnayadsktngstctqknq---qkemdrgyllgy	57						
QY	52	YPSDLNFQAPMWTSSYTGDLSPSSLELN--IPSENOYFQSAIWSGFIKVKKSDEYTPA	109						
DB	58	yfkgykdf-snlmfaptrdstllydqgtankltdkkgqgysirwlyglsksetgdfth	116						
QY	110	TSADNRHVTMWDQEVINKASNSNKRIRLEKGRLOYKIOYQENPEKGLD-----FKL	163						
DB	117	lseedegallieingklismngkekgyvnhlekylvpikleygsd--tkfidskktfkxl	174						
QY	164	YMTDSQNKKEVIVSSDNLQLEPELKOKS-----NSRKKRSTAGPTVPDRND	210						
DB	175	fklidqngqgqvgdelirmpetfnkkesgeflakpsklnlftqkmkreided---ldtdgd	231						
QY	211	GIPDSLEVEGTVVQVKNKRTFLSPWISNTHKKGGLTKYKSPKKSSTASDPYSDEKVTG	270						
DB	232	sipdlweengylt---qnrlavkwdsj--askgytkfvsnpleshtvgdpytdyekaar	286						
QY	271	RIDKNVSPARHPVLAAYRIVHVDENITLSKNEDOSTQNTSERTISKNTSTSTHTS	330						
DB	287	dlidlnaketfnpIvaafpsvnmekvllspnenls-----	323						
QY	331	EVHGNAEVHANTSTSRHTSEVHGNAEVH-----AVAIHDSLSLAGEPTWAEWTG--	380						
DB	324	---nsveshstswytlntegagaveaglgpkglstfgsvnyqhselvage--wgtstgnt	378						
QY	381	--LNTADTARLANIRYVNTGTAPITVNLPTSLVYGKQOTLATIKAKNQSLQILAPNN	438						
DB	379	sqgfinaesagylaanrynmvgtcgaaydvkptstfvl--mdtlatlataknsfahlsipge	437						
QY	439	YPSKNIAPIALNADDFESTPTMTMYNQFLFELEKTKOLRITDQOYGIATYINPENGAV	498						
DB	438	syppkqvgqgjaltsmdfnsphltlnkkkqvahllnnkpmletnqtdg---vykikdting	494						
QY	499	RVDTSNMSSEVLPOIETARTIIFNGKDLNLVERIRIAAVNPSDPLETTRPDMTLKEALKI	558						
DB	495	nltvgewngvylgqlkaktasllvdgge--rvaekrvaakdyenpedkt--psltlkdaikl	552						
QY	559	AF--GFNENGNLQYQGRKITEFDF--NPDQOTSONIKQNLDEL-----NATNITYVLDK	609						
DB	553	syppdelkelegllykxknpIyessvmtlyidentakevtqjndtgkfkdvshlydv---	609						
QY	610	IKLNKAMNILLRDKRPHYDRNNIIVAGADESVYKFEAHRVINSSTEG-----TL	657						
DB	610	-kltpkmvltk-lsllydn-----aesndnsigkvtchtnlvsgmgngkkqysnnpdanlt	664						
QY	658	LN-----	659						
DB	665	lntdagelknrdyaislysmkseknqtceltidgeilyttctktnvnhkndykrldliah	724						
QY	660	-----IDKQIRKLISGIYIEIEPTBGL	681						
DB	725	nksnpiselnhtkntdeltlfwddisltcvasikpenltdselkqlyrsygiikled--91	782						
QY	682	-----KEVIND-----RY-----DMLNITSSLRQDQKTRPI	705						
DB	783	ltdkkggylhygefineasfnlneipngnyvkcyevtsselselgnvstdlesdklykdgcltkf	842						
QY	706	DEKKY--NDKLPVYISNPNYKVNYAVT	731						
DB	843	dftkyskneqglfydsglnwdfnlnait	870						

RESULT 10

AAAI19509 standard; Protein; 884 AA.

AAAI19509;

DT	15-APR-1998	(first entry)
XX		
DE	B. cereus VIP1A(a) protein sequence.	
KW	Vegetative insecticidal protein; Bacillus cereus strain AB78; plant;	
KW	insect; Sesamia nonagrioides; maize; corn borer; toxin.	
XX		
OS	Bacillus cereus.	
PN	W09726339-A1.	
PD		
XX	24-JUL-1997.	
PF	23-DEC-1996; 96WO-EP05828.	
XX		
PR	15-JAN-1996; 96GB-0000786.	
PA	(NOVS) NOVARTIS AG.	
XX		
PI	Gay PB;	
XX		
DR	WPI; 1997-385342/35.	
XX	DR N-PSDB; AAT733994.	
PT	Protecting plants against insects of the genus Sesamia using	
PT	Bacillus toxic proteins - applied directly or expressed as	
PR	heterologous protein by the plant, also transgenic plants expressing	
XX	both Cry and VIP type toxins	
PS	Claim 8; Page 35-39; 168pp; English.	
XX		
CC	This is the amino acid sequence of the 100 kD vegetative insecticidal	
CC	protein (VIP) 1A(a) from Bacillus cereus strain AB78. The protein can	
CC	be used in a new method for protecting plants, and their progeny, against	
CC	insects of the genus Sesamia by direct or indirect application to the	
CC	plant (or seed or growing area). The protein is especially useful to	
CC	protect maize plants against the Mediterranean corn borer	
CC	(S. nonagrioides).	
SQ	Sequence 884 AA;	
QY	Query Match	20.8%; Score 783.5; DB 18; Length 884;
DB	Best Local Similarity	29.6%; Pred. No. 3.5e-44;
Matches	234; Conservative	133; Mismatches 293; Indels 131; Gaps 29;
QY	1	MKKRKVLIPLMALSTILVSS--TGNLEVIQAEVK-----OENRLNSESSESSOGILGY
DB	1	mkmmkkkiasvvtctcllapmfingvnayadsktngstctqknq---qkemdrgyllgy
QY	52	YPSDLNFQAPMWTSSYTGDLSPSSLELN--IPSENOYFQSAIWSGFIKVKKSDEYTPA
DB	58	yfkgykdf-snlmfaptrdstllydqgtankltdkkgqgysirwlyglsksetgdfth
QY	110	TSADNRHVTMWDQEVINKASNSNKRIRLEKGRLOYKIOYQENPEKGLD-----FKL
DB	117	lseedegallieingklismngkekgyvnhlekylvpikleygsd--tkfidskktfkxl
QY	164	YMTDSQNKKEVIVSSDNLQLEPELKOKS-----NSRKKRSTAGPTVPDRND
DB	175	fklidqngqgqvgdelirmpetfnkkesgeflakpsklnlftqkmkreided---ldtdgd
QY	211	GIPDSLEVEGTVVQVKNKRTFLSPWISNTHKKGGLTKYKSPKKSSTASDPYSDEKVTG
DB	232	sipdlweengylt---qnrlavkwdsj--askgytkfvsnpleshtvgdpytdyekaar
QY	271	RIDKNVSPARHPVLAAYRIVHVDENITLSKNEDOSTQNTSERTISKNTSTSTHTS
DB	287	dlidlnaketfnpIvaafpsvnmekvllspnenls-----
QY	331	EVHGNAEVHANTSTSRHTSEVHGNAEVH-----AVAIHDSLSLAGEPTWAEWTG--
DB	324	---nsveshstswytlntegagaveaglgpkglstfgsvnyqhselvage--wgtstgnt

PT both Cry and VIP type toxins
XX
PS Claim 8; Page 69-75; 168pp; English
PS

This is the amino acid sequence of a fusion protein comprising the vegetative insecticidal proteins (VIP) I_{1A} and VIP2A_{1b} from *Bacillus cereus* strain A878. The fusion protein can be used in a new method for protecting plants, and their progeny, against insects of the genus *Scemaia* by direct or indirect application to the plant (or seed or growing area). The protein is especially useful to protect maize plants against the Mediterranean corn borer (*S. nonagroides*).

SQ Sequence 1346 AA;

Query Match	20.8%	Score 783.5	DB 18	Length 1346
Best Local Similarity	29.6%	Pred. No. 6.5e-44		
Matches 234	Conservative 133	Mismatches 293	Indels 131	Gaps 29

```

OY 1 MKKRRVLPLMLASTLIVSS--TGNLEVIQAVK-----QENRLNSESSESSOLLEY 51
Db 463 mkmkmkklasvvtctllapmfmgvnaayadsktngdstctqng---qkdmrkyglly 519
OY 52 YFSDLNFOAPMVVYSTTGDISIPSELEN--IPSENOYFOASINSGFIRVKSDEYRA 109
Db 520 yfkgydf-snlmfaptrdstcllygqgtankllkkgqeygslyrwlglqsksetgdfth 578
OY 110 TSADNHVTMMVDDQEVINKASNSNKRIRLEKGRLYOIKIOYORENPTKGLD----FKL 163
Db 579 lsedegallielngklisnkgkqevahlekglylpikliegysd--tkfnidsktfkelkl 636
OY 164 YWTSQONKKEVISDNLOLPELKQSS-----NSRKRKSTAGPVPVROND 210
Db 637 fklidqnpqpgqgdelpnpefnkkesgeflakpsklnlftqkmrleld---ltldgd 693
OY 211 GIPDSELEVEGYVVDVKNKRFTFSPMYSINHEKKGLTKYKSPBEKNSTASDPYSPDEKYG 270
Db 694 sipdliweugyrl-----qurlavkxddsl-askgytkfvsnpleshtvgydpydyekaar 748
OY 271 RIDKNVSEPARHPLVAAPYIVHVDENIILSKNEDOSTONTDEFTRTISKNTSTSRTHS 330
Db 749 dlldlsnaketfnplvaafsvnvsmekvllspnehs----- 785
OY 331 EYHGAELVHANTSTGRHTSEVHGAELVH-----AVADHSLSLAGERTWALMG-- 380
Db 786 ---asveshtsenwytlnltegaseagilpbrklisfgvsvnyghsetvae--wgtslgt 840
OY 381 --LNFADTARLANANRYNVTGAPRYNYVPTSLVYCKNOGLATIKAKENOLSQLIAPNN 438
Db 841 sgfnlaasgylaanvrynmvgatgaldvkrpctsfyl-mndlatlaktaknsstalnspre 899
OY 439 YPSPKNLAPILNAOADESSPTPTNNYNOFLLELETKOQLRDTDOYGNIAIYFNENGVR 498
Db 900 syprkgqgqgialtsmddnshpirlnkkqvdnllmkkpmmlietnqtdg---vykkltdng 956
OY 499 RVDGSSNMSEVLPOIETTARIIFNGKDLNLVERIAAIVNPSDPLETTKPDMLTKALKI 558
Db 957 nlvtgwmngvlyqkaktasiallvddge-rvaeakrvaaqdyenpedkt-psltlkcalkl 1014
OY 559 AF--GFENPNGLUQOGKADITEFD--NPDQOTSNIKQLAEL-----NAINIYIVLDK 609
Db 1015 syrpdelkelegllyknpriyesswmtlyldenlakvcltqldtqgkfkdvshlydv--- 1071
OY 610 IKLNKKNMILIDKRFHFDNRNIAVGADESIVYKAEHREINSGTES-----TL 657
Db 1072 -klprkmvntlk-isllydn---aesnduslqkwtlnltvsgngngkkkyssnnpaant 1126
OY 658 LNID-----KDIRKILSGYIVEIEDTEGLKEVINDRYDMLNIS--SLRDGGTFIDFKKY 710
Db 1127 lntdageklmrkrdyrlslymkseknqtgcetlidgeilgyipcltktyvnvndknykrlidiah 1186
OY 711 NDKLPLVLSNP 721

```

ID	AAW46723	standard: Protein; 1346 AA.
XX	AAW46723;	
XX	02-JUN-1998	(first entry)
XX	VIP1A(a)/VIP2A(a)	fusion protein.
XX	Vegetative insecticidal protein; VIP; expression; maize; protection;	
KW	plant; <i>Ostrinia furnacalis</i> ; Asian Corn Borer; Cry toxin; VIP toxin;	
KM	recombinant; <i>Bacillus thuringiensis</i> ; transgenic plant; resistance;	
KW	insect attack; <i>Sesamla</i> ; maize; cereal crop.	
XX	Synthetic.	
OS	<i>Bacillus</i> sp.	
PN	WO9746105-A1	
XX	11-DEC-1997.	
PD	27-MAY-1997;	97WO-EP02737.
PF	06-JUN-1996;	96GB-0011777.
XX	(NOVS)	NOVARTIS AG.
PA	Hunter B,	Suwanataradon K, Utdewilligen WPM;
PI	WPI; 1998-041787/04.	
XX	N-PSDB; AAV16172.	
DR	Administration of <i>Bacillus</i> sp. toxin protein, especially Cry or	
PT	vegetative insecticidal protein (VIP) protein to plants - useful for	
PT	protection against attack by Asian Corn Borer (<i>Ostrinia furnacalis</i>)	
PS	Claim 8; Pages 78-83; 175pp; English.	
XX	The present sequence represents the fusion protein of vegetative	
CC	insecticidal protein 1A(a) (VIP1A(a)) and VIP2A(a). The protein is	
CC	used in a method for protecting plants and their progeny against	
CC	damage caused by <i>Ostrinia furnacalis</i> (Asian Corn Borer). The	
CC	protein is directly or indirectly applied to the plant, plant seed or	
CC	growing area of the plant. Cry toxins can also be used in the same way,	
CC	in place of VIP toxins. The Cry or VIP toxins and genes are used,	
CC	especially inside recombinant <i>B. cereus</i> or <i>B. thuringiensis</i> strains, to	
CC	produce plants protected against Asian Borer pests. Transgenic plants	
CC	protected against Asian Corn Borer can be used to produce seed and	
CC	progeny also resistant to insect attack. Plants expressing both a	
CC	Cry-type and a VIP toxin gene can also protect against <i>Sesamia</i> pests.	
CC	The method and compositions are especially used for protecting maize	
CC	but may also be used to protect other cereal crops against Asian Corn	
CC	Borer attack.	
XX	Sequence	1346 AA;
XX	Query Match	20.8%; Score 783.5; DB 19; Length 1346;
XX	Best Local Similarity	29.6%; Pred. No. 6,5e-44;
XX	Matches 234; Conservative 133; Mismatches 293; Indels 131; Gaps	29;
OY	1 MKRRVLPALMSTLVSS--TGNLEVIOAEVK-----OENRLNSESSESGGLGCV 51	
DB	463 mkmmkkkllasvvtcllpmfngvnavyadsktngtstqknq---qkmdrkgylgy 519	
OY	52 YPSDNLFOAPAMVYSSSTGDLSPSSLELN--IPSPNOYFOALNSGFLKVKKSDERTPA 109	
DB	520 ytkgkdf-snlmftaprtcdsllydqgtanklldkkqggygsitwvlgilqsketgdtfn 578	

```

PD      11-DEC-1997.
XX
XX
XX      27-MAY-1997;      97WO-EP02737.
PE
XX
XX      06-JUN-1996;      96GB-0011777.
XX
XX      (NOVS ) NOVARTIS AG.
PA
XX
XX      Hunter B, Suwantaradon K, Utdewilligen WPM;
PI
DR      WPI: 1998-04187/04.
DR      N-PSDB; AAV16172.
XX
XX      Administration of Bacillus sp. toxin protein, especially Cry or
PT      vegetative insecticidal protein (VIP) protein to plants - useful for
PT      protection against attack by Asian Corn Borer (Ostrinia furnacalis)
PS
XX      Claim 8; Pages 78-83; 175pp; English.
PS
XX
XX      The present sequence represents the fusion protein of vegetative
CC      insecticidal protein 1A(a) (VIP1A(a)) and VIP2A(a). The protein is
CC      used in a method for protecting plants and their progeny against
CC      damage caused by Ostrinia furnacalis (Asian Corn Borer). The
CC      protein is directly or indirectly applied to the plant, plant seed or
CC      growing area of the plant. Cry toxins can also be used in the same way,
CC      in place of VIP toxins. The Cry or VIP toxins and genes are used,
CC      especially inside recombinant B. cereus or B. thuringiensis strains, to
CC      produce plants protected against Asian Borer pests. Transgenic plants
CC      protected against Asian Corn Borer can be used to produce seed and
CC      progeny also resistant to insect attack. Plants expressing both a
CC      Cry-type and a VIP toxin gene can also protect against Sesamia pests.
CC      The method and compositions are especially used for protecting maize
CC      but may also be used to protect other cereal crops against Asian Corn
CC      Borer attack.
XX
XX      Sequence      1346 AA;
SQ

```

```

OY 110 TSADNHVTWVDDQEVINKASNNKIRLEKGRLYOIKIOYQENPEKGLD-----FKL 163
DB 579 lsedegallengklisnkgkqvnhlekylpikleygsd--tkfidsktkekl 636
OY 164 WYTDSONKKEVISDNLOPELKOSS-----NSRRKSTSTAGTVPDRDND 210
DB 637 fklidqngqgvgqdelirpefnkkesgeflakpsklnlftqkmkrelded---tdtdgd 693
OY 211 GIPDSLEVEGYVVDVKNKRTPLSPWISNHEKKGLTKYKSSPEKSTASDPYSDPEKVTG 270
DB 694 slpdlweengytl-----qnrliavkwddsl-askgytkfvsnpleshtvgdpytdyekaar 748
OY 271 RIDKNVSPARRPLVAAYPIVHVDMENTILSKNEDOSTONTSERTISKNTSTSTHTS 330
DB 749 dldlshnaketlplvaafpsvvnsmekvllspnenls----- 785
OY 331 EVHGNAEVAHNTSTSTRTSTSEVHGNAEVH-----AVAIHSLSLAGERTWAETMG-- 380
DB 786 ---nveshststnwsytnlegasveaglsfgsvnyghsetvage--wgtstgnt 840
OY 381 --LNTADTARLANIRYVNTGTAPLYNVLPTTSLVLGKNQTLATIKAKENQSLILAPNN 438
DB 841 seqfntasagylaanvrynmgvgalydvkptsfvl-ndtlatlakastalnspge 899
OY 439 YPSPNNLAPIALNADDESTPTIMWYNOFLELEKTKOLRDTDOYGNATYNNENGKV 498
DB 900 syppkqngqgialtsmdtdinspdltnkkqvdnltnkpmmlenqtdg---vykklktdtg 956
OY 499 RVDTSNNSEVLPOIQTETARIIFNGKDLNLERRIAAVNSPDPLETTPTPDMTLKEALKI 558
DB 957 nlvtgsewngvqglkakastlvdgge-irvaekrvaadynpenekt-psltlkldakl 1014
OY 559 AF--GFNENGLQYOGKDIETEDF--NPDQOTSONIKQOLAE-----NATNIYTVLDK 609
DB 1015 sypldtkellegllyknpkilyesswmtlyidentakevtqlndtgkfkdvshlydv--- 1071
OY 610 ICLNKMNLINDKRFPHYDRNNIAGADESVYKEAHREVINSTES-----LL 657
DB 1072 -kltpkmvclx-lsllydn---aesndnsigkwtlnltnsgngnkkkysnnpdaant 1126
OY 658 INID-----KDIRKILSGYIIEDETEGLKEVINDRYDMLNIS-SLRDQKTFIDFKKY 710
DB 1127 lntdageklnkrdylyslismkseknltgeitldgeilyptlctktnvvnkdnkridlsh 1186
OY 711 NDKLPLYISNP 721
DB 1187 ntk-----snp 1192

```

RESULT 15

```

AAM60224
ID AAM60224 standard; Protein; 880 AA.
AC AAM60224;
XX
XX 28-SEP-1998 (first entry)
XX
XX Bacillus thuringiensis insecticidal toxin 177C8.
XX
XX Insecticide; pesticidal; toxin; delta-endotoxin;
XX biological control; lepidopteran; coleopteran.
XX
XX Bacillus thuringiensis strain PS177C8 (NRRL B-21867).
XX
XX Key Location/Qualifiers
XX FT MISC-difference 253
XX FT MISC-difference 675 /note= "encoded by YTA"
XX FT MISC-difference 846 /note= "encoded by AC"
XX FT MISC-difference 846 /note= "encoded by RAA"
XX
XX W09818932-A2.

```

```

XX 07-MAY-1998.
XX 30-OCT-1997; 97WO-US19804.
XX 30-OCT-1996; 96US-0029848.
XX (MYCO ) MYCOGEN CORP.
XX Dullum CJ, Feltelson JS, Loewer D, Muller-Cohn J;
XX Narva KE, Schmelts JL, Schneft HE, Schwab G, Stamp L;
XX Stockhoff BA;
XX WPI; 1998-272226/24.
XX N-PSDB; AAV30307.
XX
XX Bacillus thuringiensis isolates - used for producing pesticidal
XX toxins and nucleotide sequences for control of lepidopterans and
XX coleopterans
XX
XX Claim 5; Page 81-84; 139p; English.
XX
XX This polypeptide comprises a novel soluble toxin of Bacillus
XX thuringiensis (B.t.) strain PS177C8 (NRRL B-21867). The toxin
XX belongs to a novel family of B.t. toxins that have toxicity
XX against non-mammalian pests. Its amino acid sequence was deduced
XX from a novel DNA fragment (see AAV30307) obtained by PCR from
XX cellular genomic DNA of PS177C8. Disclosed and claimed are novel
XX B.t. isolates and toxins (see AAM60218-32) that have activity against
XX lepidopteran and/or coleopteran pests, isolated genes, probes
XX and primers (see AAV30288-321 and AAV99734-87) useful for production
XX of the toxins and for the identification and characterisation of
XX these toxins, and transformed hosts, particularly plant and
XX bacterial hosts. The invention provides 8 entirely new families of
XX toxins from B.t. isolates. The toxins have the additional ability
XX to form pores in cell membranes, and can be used to facilitate
XX entry of a second agent into a target cell.
XX
XX Sequence 880 AA:
XX
XX Query Match 20.8%; Score 781.5; DB 19; Length 880;
XX Best Local Similarity 29.4%; Pred. No. 4,7e-44;
XX Matches 234; Conservative 136; Mismatches 293; Indels 133; Gaps 30;
XX
XX 1 MKRRKYLPLMALSTILVSS--TGNLEVIOAEVK-----OENRLNSESSESSOGLLYG 51
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
XX 1 mkkk---lasvvtccllapmfingvnvayadskntqstgknq---qkmdrfqllgy 54
XX
XX 52 YPSDLNFOAPMVTSTTGDLSPSSLEN--IPSENOYFOSAINSGFTKVKSDPYTFA 109
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
XX 55 yfkgydf-snlmfaptrdstllydqtankltdkkgqgsirvlglsksetgdfth 113
XX
XX 110 TSADNHVTWVDDQEVINKASNNKIRLEKGRLYOIKIOYQENPEKGLD-----FKL 163
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
XX 114 lsedegallengklisnkgkqvnhlekylpikleygsd--tkfidsktkekl 171
XX
XX 164 WYTDSONKKEVISDNLOPELKOSS-----NSRRKSTSTAGTVPDRDND 210
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
XX 172 fklidqngqgvgqdelirpefnkkesgeflakpsklnlftqkmkrelded---tdtdgd 228
XX
XX 211 GIPDSLEVEGYVVDVKNKRTPLSPWISNHEKKGLTKYKSSPEKSTASDPYSDPEKVTG 270
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
XX 229 slpdlweengytl-----qnrliavkwddsl-askgytkfvsnpleshtvgdpytdyekaar 283
XX
XX 271 RIDKNVSPARRPLVAAYPIVHVDMENTILSKNEDOSTONTSERTISKNTSTSTHTS 330
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
XX 284 dldlshnaketlplvaafpsvvnsmekvllspnenls----- 320
XX
XX 331 EVHGNAEVAHNTSTSTRTSTSEVHGNAEVH-----AVAIHSLSLAGERTWAETMG-- 380
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
XX 321 ---nveshststnwsytnlegasveaglsfgsvnyghsetvage--wgtstgnt 375

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:55:15 ; Search time 132.69 seconds
(without alignments) 615.118 Million cell updates/sec

Title: US-09-747-521-4_COPY_178_735

Perfect score: 2871
Sequence: 1 DNQLQPELKKQSSNSKKRRS.....LYISNPYKVNYATYKENT 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.potent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2755	96.0	764	2	09R0U2
2	2751	95.8	764	2	09KH69
3	2743	95.5	764	2	09F5R7
4	729	25.4	876	2	09KH41
5	728	25.4	876	2	032739
6	718.5	25.0	879	2	006498
7	715.5	24.9	875	2	046221
8	643.5	22.4	721	2	086171
9	211.5	7.4	204	2	09X377
10	181.5	6.3	2178	2	046149
11	166	5.8	604	5	026021
12	164.5	5.7	2529	2	025579
13	158	5.5	4688	2	09P008
14	157	5.5	1308	5	096129
15	155.5	5.4	559	5	09U3Y8
16	150.5	5.2	2269	5	026223
17	150.5	5.2	2399	2	09ZKS9
18	150	5.2	1127	12	09YVT6
19	150	5.2	4919	2	09ZHU0

20	149.5	5.2	2747	5	09BUX9	09bux9 plasmidium
21	149	5.2	1193	2	P71107	P71107 clostridium
22	149	5.2	1193	2	045914	045914 clostridium
23	148.5	5.2	622	2	09U0K4	09u0k4 plasmidium
24	148	5.2	810	2	09CFD3	09cfcd3 lactococcus
25	147	5.1	1156	5	077317	077317 plasmidium
26	147	5.1	1939	5	025662	025662 plasmidium
27	146.5	5.1	655	5	026109	026109 plasmidium
28	146.5	5.1	3978	5	097236	097236 plasmidium
29	145	5.1	2647	5	09U4X0	09u4x0 plasmidium
30	144.5	5.0	7107	5	09VAF7	09vaf7 drosophila
31	144	5.0	2402	2	09AER7	09aer7 staphylococ
32	144	5.0	2522	5	077365	077365 plasmidium
33	143.5	5.0	1786	5	09U0P0	09u0p0 plasmidium
34	143	5.0	2340	2	09ZD91	09zd91 rickettsia
35	142.5	5.0	1072	2	09CF64	09cf64 lactococcus
36	142.5	5.0	6658	5	076281	076281 drosophila
37	142	4.9	1595	2	052373	052373 caldicellul
38	141.5	4.9	1893	5	09NKC9	09nkc9 drosophila
39	141.5	4.9	1928	5	09U0H2	09u0h2 plasmidium
40	141.5	4.9	4152	2	09ZHU3	09zhu3 haemophilus
41	141	4.9	712	13	09S8N5	09s8n5 gallus gall
42	141	4.9	713	2	09CM95	09cm95 pasteurilla
43	141	4.9	786	2	09FDM5	09fdm5 streptococ
44	141	4.9	2771	5	026216	026216 plasmidium
45	140.5	4.9	1557	2	09RNI2	09rni2 haemophilus

ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	764 AA.
09R0U2	09R0U2			
AC	09R0U2:			
DP	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	PX01-110 (PROTECTIVE ANTIGEN).			
GN	PAG.			
OS	Bacillus anthracis.			
OG	Plasmid virulence plasmid PX01, and plasmid PX01.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
CC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1392;			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=STERNE; PLASMID=VIRULENCE PLASMID PX01;			
RC	STRAIN=99445483; PubMed=10515943;			
RA	Okimaka R.T., Cloud K., Hampton O., Hoffmaster A.R., Hill K.K.,			
RA	Keim P., Koehler T.M., Lamke G., Kumano S., Mallion J., Menter D.,			
RA	Marlinez Y., Ricker D., Svensson R., Jackson P.J.,			
RT	"Sequence and organization of px01, the large Bacillus anthracis			
RT	plasmid harboring the Anthrax toxin genes.";			
RL	J. Bacteriol. 181:6509-6515(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=33, 28, AND BA1035; PLASMID=PX01;			
RC	STRAIN=99214082; PubMed=10197996;			
RX	Price L.B., Hugh-Jones M., Jackson P.J., Keim P.,			
RA	"Genetic diversity in the protective antigen gene of Bacillus			
RT	anthracis.";			
RL	J. Bacteriol. 181:2358-2362(1999).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=33, 28, AND BA1035; PLASMID=PX01;			
RC	Price L.B., Hugh-Jones M., Jackson P.J., Keim P.,			
RA	Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AF065404; AAD32414.1; -			
DR	EMBL; AF306781; AAG24449.1; -			
DR	EMBL; AF306778; AAG24446.1; -			
DR	EMBL; AF306779; AAG24447.1; -			
DR	EMBL; AF306780; AAG24448.1; -			

DR HSP: P13423; JACC.
DR InterPro: IPR003896; Binary_toxB.
DR PRINTS: PRO1391; BINARYTOXINB.
KW Plasmid.
SQ SEQUENCE 764 AA; 85810 MW; 3AE1EFBF48FAA03F CRC64;

Query Match 96.0%; Score 2755; DB 2; Length 764;
Best Local Similarity 96.6%; Pred. No. 2, 9e-157;
Matches 539; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

QY 1 DNLQPELKOKSSNRKRSKSTAGPTVPDRDNDGIPDSLEVGTYVDVKNKRTPLSPWIS 60
DB 178 DNLQPELKOKSSNRKRSKSTAGPTVPDRDNDGIPDSLEVGTYVDVKNKRTPLSPWIS 237
QY 61 NIHEKGLTKYKSSPEKWSSTADSPYSDPEKVTGRIDKNVSPKARHPLVAAYPIVHVDMEN 120
DB 238 NIHEKGLTKYKSSPEKWSSTADSPYSDPEKVTGRIDKNVSPKARHPLVAAYPIVHVDMEN 297
QY 121 ILSKNEOSTONTSETRTISKNTSTRTHTSEVGNAEVIANSTSTRTHTSEVGNAE 180
DB 298 ILSKNEOSTONTSETRTISKNTSTRTHTSEVGNAEVIANSTSTRTHTSEVGNAE 357
QY 181 VHAVALDHSLSLAGERTAETMGLNADTARLANIRYVNTGAPLYNVLPPTSLVLGN 240
DB 358 SSTVALDHSLSLAGERTAETMGLNADTARLANIRYVNTGAPLYNVLPPTSLVLGN 417
QY 241 QTLATIKAKENOLSOILAPNNYPSKNLAPIALMAODDFSSPTITMNNYNOFLELEKTOL 300
DB 418 QTLATIKAKENOLSOILAPNNYPSKNLAPIALMAODDFSSPTITMNNYNOFLELEKTOL 477
QY 301 RLDTOOVGNITATYNEENGRVVDGSMNSEVLPQIOETTARIIFNGKDLNVERRIAY 360
DB 478 RLDTOOVGNITATYNEENGRVVDGSMNSEVLPQIOETTARIIFNGKDLNVERRIAY 537
QY 361 NPSDPLETTKPPMTLEKALIAFGFNEPNGNLOYOGKDTTEPDFNDQOTSONIKQOLAE 420
DB 538 NPSDPLETTKPPMTLEKALIAFGFNEPNGNLOYOGKDTTEPDFNDQOTSONIKQOLAE 597
QY 421 LNATNIYTVLDRIKLNKAKNNILIRDRFRHYDRNNIAVGADESVEKAEHREVINSTEGLL 480
DB 598 LNATNIYTVLDRIKLNKAKNNILIRDRFRHYDRNNIAVGADESVEKAEHREVINSTEGLL 657
QY 481 LNIIDDIRKILSGYIEIDTEGLKEVINDRYDMLNISSLRDQGTFFIDFKKYNDKPLX 540
DB 658 LNIIDDIRKILSGYIEIDTEGLKEVINDRYDMLNISSLRDQGTFFIDFKKYNDKPLX 717
QY 541 ISNPYKVVAVYATKENT 558
DB 718 ISNPYKVVAVYATKENT 735

RESULT 2
Q9KH69 PRELIMINARY: PRT: 764 AA.
AC Q9KH69:
DB 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE PROTECTIVE ANTIGEN.
GN PACG OR PAG.
OS Bacillus anthracis.
OG Plasmid pX01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V770-NP1-R, ATCC14185; PLASMID=PX01;
RX MEDLINE=20359347; PubMed=10899854;
RA Cohen S., Mendelson I., Alboun Z., Koblner D., Elhanany E., Bino T.,
RA Lettner M., Inbar I., Rosenberg H., Gozes Y., Barak R., Fisher M.,
RA Kromann C., Velan B., Shaffer A.

RT "Attenuated nontoxicogenic and nonencapsulated recombinant Bacillus
RT anthracis spore vaccines protect against anthrax".
RL Infect. Immun. 68:4549-4558(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC PLASMID=PX01;
RX MEDLINE=99214082; PubMed=10197996;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RT "Genetic diversity in the protective antigen gene of Bacillus
RT anthracis".
RL J. Bacteriol. 181:2358-2362(1999).

RN [3]
RP SEQUENCE FROM N.A.
RC PLASMID=PX01;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF268967; AAF6457.1; -
DR EMBL: AF306782; AAG24450.1; -
DR InterPro: IPR003896; Binary_toxB.
DR PRINTS: PRO1391; BINARYTOXINB.
KW Plasmid.
SQ SEQUENCE 764 AA; 85838 MW; B0DAFC1DCF83DAF4 CRC64;

Query Match 95.8%; Score 2751; DB 2; Length 764;
Best Local Similarity 96.4%; Pred. No. 5, 1e-157;
Matches 538; Conservative 3; Mismatches 17; Indels 0; Gaps 0;

QY 1 DNLQPELKOKSSNRKRSKSTAGPTVPDRDNDGIPDSLEVGTYVDVKNKRTPLSPWIS 60
DB 178 DNLQPELKOKSSNRKRSKSTAGPTVPDRDNDGIPDSLEVGTYVDVKNKRTPLSPWIS 237
QY 61 NIHEKGLTKYKSSPEKWSSTADSPYSDPEKVTGRIDKNVSPKARHPLVAAYPIVHVDMEN 120
DB 238 NIHEKGLTKYKSSPEKWSSTADSPYSDPEKVTGRIDKNVSPKARHPLVAAYPIVHVDMEN 297
QY 121 ILSKNEOSTONTSETRTISKNTSTRTHTSEVGNAEVIANSTSTRTHTSEVGNAE 180
DB 298 ILSKNEOSTONTSETRTISKNTSTRTHTSEVGNAEVIANSTSTRTHTSEVGNAE 357
QY 181 VHAVALDHSLSLAGERTAETMGLNADTARLANIRYVNTGAPLYNVLPPTSLVLGN 240
DB 358 SSTVALDHSLSLAGERTAETMGLNADTARLANIRYVNTGAPLYNVLPPTSLVLGN 417
QY 241 QTLATIKAKENOLSOILAPNNYPSKNLAPIALMAODDFSSPTITMNNYNOFLELEKTOL 300
DB 418 QTLATIKAKENOLSOILAPNNYPSKNLAPIALMAODDFSSPTITMNNYNOFLELEKTOL 477
QY 301 RLDTOOVGNITATYNEENGRVVDGSMNSEVLPQIOETTARIIFNGKDLNVERRIAY 360
DB 478 RLDTOOVGNITATYNEENGRVVDGSMNSEVLPQIOETTARIIFNGKDLNVERRIAY 537
QY 361 NPSDPLETTKPPMTLEKALIAFGFNEPNGNLOYOGKDTTEPDFNDQOTSONIKQOLAE 420
DB 538 NPSDPLETTKPPMTLEKALIAFGFNEPNGNLOYOGKDTTEPDFNDQOTSONIKQOLAE 597
QY 421 LNATNIYTVLDRIKLNKAKNNILIRDRFRHYDRNNIAVGADESVEKAEHREVINSTEGLL 480
DB 598 LNATNIYTVLDRIKLNKAKNNILIRDRFRHYDRNNIAVGADESVEKAEHREVINSTEGLL 657
QY 481 LNIIDDIRKILSGYIEIDTEGLKEVINDRYDMLNISSLRDQGTFFIDFKKYNDKPLX 540
DB 658 LNIIDDIRKILSGYIEIDTEGLKEVINDRYDMLNISSLRDQGTFFIDFKKYNDKPLX 717
QY 541 ISNPYKVVAVYATKENT 558
DB 718 ISNPYKVVAVYATKENT 735

RESULT 3
Q9F5R7 PRELIMINARY: PRT: 764 AA.
AC Q9F5R7:
DB 09F5R7:

DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PROTECTIVE ANTIGEN.
GN Bacillus anthracis.
OS Bacillus anthracis.
OC Plasmid pX01.
OC Bacteria: Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-BAL024;
MEDLINE=99214082; PubMed=10197996;
RA Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RT "Genetic diversity in the protective antigen gene of Bacillus anthracis";
RL J. Bacteriol. 181:2358-2362(1999).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN-BAL024;
Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF306783; AAG24451.1; -
DR InterPro: IPR003896; Binary toxB.
DR PRINTS: PR01391; BINARYTOXINB.
KW Plasmid.
SQ SEQUENCE 764 AA; 85828 MW; A1845CE1FEDCD93A CRC64;

Query Match 95.5%; Score 2743; DB 2; Length 764;
Best Local Similarity 96.2%; Pred. No.1.5e-156;
Matches 537; Conservative 3; Mismatches 18; Indels 0; Gaps 0;

OY 1 DNLQPELKQSSNSRRKSTSGAPTPVDRDNDGIPDSLEEGYTVDKKRFPLSPWIS 60
DB 178 DNLQPELKQSSNSRRKSTSGAPTPVDRDNDGIPDSLEEGYTVDKKRFPLSPWIS 237
OY 61 NIEHKKGLTYKSSPEKWSASDPYSDFEYVGRIDKNVSPKAPLVAAPYIHYVMEN 120
DB 238 NIEHKKGLTYKSSPEKWSASDPYSDFEYVGRIDKNVSPKAPLVAAPYIHYVMEN 297
OY 121 IILSKNEDOSTONTDESETRTSKNTSTSRTHSEVHGNAEYHANTSTSRTHSEVHGNAE 180
DB 298 IILSKNEDOSTONTDESETRTSKNTSTSRTHSEVHGNAEYHANTSTSRTHSEVHGNAE 357
OY 181 VHAVALDHSLSLAGESTWAGTGLNTADTARLANANIRYVNTGAPYIYNVLPPTSLSVLGKN 240
DB 358 SSVYVAIDHSLSLAGESTWAGTGLNTADTARLANANIRYVNTGAPYIYNVLPPTSLSVLGKN 417
OY 241 QTLATITAKENQSLQILAPNNYPSKNLAPLALNAODFSSPTITMNYNOFLELEKTKOL 300
DB 418 QTLATITAKENQSLQILAPNNYPSKNLAPLALNAODFSSPTITMNYNOFLELEKTKOL 477
OY 301 RLPTDQYGGIATYNFENGVRVDTGSMSEVLPOIQTETARIIFNGKDLVVERRTAAV 360
DB 478 RLPTDQYGGIATYNFENGVRVDTGSMSEVLPOIQTETARIIFNGKDLVVERRTAAV 537
OY 361 NPSDPLETTKPDMTLKEALKIAFGFENPNCNLOYOGKDIETFDENFDQOSTONIKNOAE 420
DB 538 NPSDPLETTKPDMTLKEALKIAFGFENPNCNLOYOGKDIETFDENFDQOSTONIKNOAE 597
OY 421 LANTNTITVTDKIKLANKNMILIRDKRFHYDRNNIAVGADESIVKKAHREYINSTGGL 480
DB 598 LANTNTITVTDKIKLANKNMILIRDKRFHYDRNNIAVGADESIVKKAHREYINSTGGL 657
OY 481 LNTIDKDIRKILSGYIEIEPTGKLEVINDRYDMLNITSSRODGKTFIDFKKYNDKPLX 540
DB 658 LNTIDKDIRKILSGYIEIEPTGKLEVINDRYDMLNITSSRODGKTFIDFKKYNDKPLX 717
OY 541 ISNPNKYVNVYAVTKENT 558
DB 718 ISNPNKYVNVYAVTKENT 735

RESULT 4
O9KH41 PRELIMINARY; PRT; 876 AA.
AC O9KH41;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CDTB.
GN CDTB.
OS Clostridium difficile.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium
OX NCBI_TaxID=1496;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CCUG 20309;
RA Chang S.Y., Song K.P.;
RT "ADP-ribosylating Binary Toxin Genes of Clostridium difficile strain CCUG 20309";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF271719; AAF81761.1; -
DR InterPro: IPR003896; Binary toxB.
DR PRINTS: PR01391; BINARYTOXINB.
SQ SEQUENCE 876 AA; 98792 MW; 360D62F352E745A5 CRC64;

Query Match 25.4%; Score 729; DB 2; Length 876;
Best Local Similarity 35.3%; Pred. No.1.1e-35;
Matches 200; Conservative 95; Mismatches 194; Indels 78; Gaps 22;

OY 29 DRDNDGIPDSLEEGYTVDKKRFPLSPWISNIEHKKGLTYKSSPEKWSASDPYSDF 88
DB 220 DTDNDGIPDSLEEGYTVDKKRFPLSPWISNIEHKKGLTYKSSPEKWSASDPYSDF 274
OY 89 EKYVGRIDKNVSPKAPLVAAPYIHYVMENIILSKNEDOSTONTDESETRTSKNTS 148
DB 275 EKASGFDKATKTERADPLVAAPYIVGMEKLIITNEHAST----DQKTVSRATINS 330
OY 149 RTHTSEVHGNAEY-----HANTSTSRTHSEVHGNAEYHANTSTSRTHSEVHGNAEY 202
DB 331 KTESNFGVSVNVGYONGFTANVTNVSHTD-----NSTAVQDS--NGE-SWMTGL 379
OY 203 GLNTADTARLANANIRYVNTGAPYIYNVLPPTSLSVLGKNQTLATITAKENQSLQILAPNNY 262
DB 380 SINGESAYIANANRYVNTGAPYIYNVLPPTSLSVLGKNQTLATITAKENQSLQILAPNNY 438
OY 263 YPSKNLAPLALNAODFSSPTITMNYNOFLELEKTKOLRLPTDQYGGIATYNFENGVRV 322
DB 439 YPKGGLSPALANTMDQSSRLPIPTNIDQLKLDGKQIKLETTOVSGFGTKN-SSQOI- 496
OY 323 VDTGSMSEVLPOIQTETARIIFNGKDLVVERRTAAVNSDPLETTKPDMTLKEALKIA 382
DB 497 VTEGNSMSDYISQIDTSASITLIDTEN-ESYERVRTAKNLQDPEDDKI-PELTIGEAIEKA 554
OY 383 FGFNEPNCNLOYOGKDIETFDENFDQOSTONIKNOAEIANTNTITVTDKIKLANKN 440
DB 555 FGAATKKGGLYFNDIPIDESCVELIFDNDTANKIKDSIKTISDKKIYV--KLEGMN 610
OY 441 ILIDKRFHY-----DRNNIANGADESVKKAHREYINSTGGL-----LNTIDDIR 488
DB 611 ILITPTPTFTNFDYNNP--STWSNVNTNKKDGLQGSANKLNGEYIKIIPMSLEKPYKR 668
OY 489 KILSGY-----IVEI-----EDTEGL-----KEVINDRYDMLNIS-SLRQD 523
DB 669 YVFGYSGKDPFTSIIIVKIKAKKEKTDYLVPEQGYTKFVSFEPTTEKSSNIEFTLIGS 728
OY 524 GKTFIDFKKYND--KLPLYISNPNKYV 548
DB 729 GTTYLDMLSITELNSTPEIIDEPEVKI 755

RESULT	5			
032739				
ID	032739	PRELIMINARY;	PRT;	876 AA.
AC	032739;			
DT	01-JAN-1998 (TREMBLrel. 05, Created)			
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	ADP-RIBOSYLTRANSFERASE.			
GN	CDPB.			
OS	Clostridium difficile.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_Taxid=1496;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CD196;			
RX	MEDLINE=97230316; PubMed=9119480;			
RA	Perelle S., Gilbert M., Bourlioux P., Corthier G., Popoff M.R.;			
RT	"Production of a complete binary toxin (actin-specific ADP-			
RL	ribosyltransferase) by Clostridium difficile CD196.";			
DR	Infect. Immun. 65:1402-1407(1997).			
DR	EMBL; L76081; AAB67305.1; -.			
DR	HSSP; P13423; IACC.			
DR	InterPro; IPR003896; Binary_toxB.			
DR	PRINTS; PRO1391; BINARYTOXINB.			
KW	Transferase			
SQ	SEQUENCE 876 AA: 98796 MW; 25E06E2D45CE2B3B CRC64;			

Query Match	25.4%	Score 728	DB 2:	length 876;
Best Local Similarity	35.3%	Pred. No. 1	2e-35:	
Matches 200;	Conservative 95;	Mismatches 194;	Indels 78;	Gaps 22

QY	29	DRDNDGIDPSLEVEEYETADVKKRPFELSPWISNIEHKQITLTKKSPKEMKXASPDYOSOF	88
Db	220	DTDNDNDIPDSIERNCYTTI-----KDLIAVKWEDSEAE-QCYKRYVSNYLESNATGADPIDY	274
QY	89	EKYVGRIDKXNSPEARHBLVAAPYIVHVDMENILISKNEDOSOTONTDEPRTISKNTSTS	148
Db	275	EKASGSPDKAIKTEARDEPLVAAPYIVGOMEXMLISTNEHAST-----DQKIVSRATNTS	330
QY	149	RTHSEVGNAAVEY-----HANTSTSRHNTSEVGHGNAEYHAAIDHSLSLAGEPMAETM	202
Db	331	KTESSTAGVSXVWGONGETANVTNTYSHHTD-----NSTAYDOS---NGE-SWNTGL	379
QY	203	GLNTPDTRLANANIKRYVWTGAPYINVLPTTSLVYLGKNQOTLTIKAKENQSLQIILANPNY	262
Db	380	SINKGESATYINANVYTYTGTPAKKYVPTTLVL--DGDTLSTIAEOENOJNNLSPGOT	438
QY	263	YPSKNLADIALNAODFSSPTITMKNYNOFLELEKTRKOLRLDYOYVGNATATYFENGVR	322
Db	439	YPKKGLSPLALNTMGOFFSRLPIPVYDQKRLDACKQKILETTQVSGNNGTAN--SSGQI-	496
QY	323	VDTGSNMBSEVLRPOIOETTARITIFNGCKDLNLYRRIRLAAYVSPPLETTKPKDMTLKALKTA	382
Db	497	VTEGMSWSDYSQIDISIASIILDTEN-ESYERKRYTAKNLOPEKPT--PELTIGEAIEKA	554
QY	383	FGFNPNPNLLOYQGDIRE--FDNPEDOOTSINIKNOELNATNITYTLDKIKKANAKN	440
Db	555	FGATRKDOLLFNNDIPIDESCYELLFDONTAKIKDLSLTKLTSODKRIYV---KLERGMN	610
QY	441	ILIRDKRPHY--DRNNIATVGADESVEAREHVNINSSTEGL-----LLNIDKDIR	488
Db	611	ILIKRPYFTNDDOYNNP--STWGNVMTNTQDGLGSGANKLNGETKIKIIPSELKPYKR	668
QY	489	KILSGY-----IYEI-----EDTEGL-----KVIINDRYDMLNIS--SLROD	523
Db	669	YVFSGYSKDPLTNSIIVYIKAKKEKRTDYLVPEOGYKFSYEFETTEKXDSNMEITLLGS	728
QY	524	GKTFIDFKKYN--KLPLYISNPNKV	548
Db	729	GTTYLDNLSITELNSTPELLIDEPEVKI	755

RESULT	6			
006498				
ID	006498	PRELIMINARY:	PRT:	879 AA.
AC	006498;			
DT	01-JUL-1997 (TREMBLrel. 04, Created)			
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	SB COMPONENT.			
GN	SBS.			
OS	Clostridium spiroforme.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes.			
OX	NCBI_TaxID=29348;			
BN	[1]			
RP	SEQUENCE FROM N.A.			
EC	STRAIN=CS246;			
EA	Gilbert M., Perelle S., Daube G., Popoff M.R.;			
EL	Syst. Appl. Microbiol. 20:337-347(1997).			
DR	EMBL; X97969; CAA6612.1; -;			
DR	HSSP; P13423; IAC.			
DR	InterPro; IPR003896; Binary_toxB.			
DR	PRINTS; PRO1391; BINARYTOXINB.			
SO	SEQUENCE 879 AA; 98738 MW; 40685ACBBE05BA01 CRC64;			

Query Match	25.0%;	Score 718.5;	DB 2;	Length 879;
Best Local Similarity	29.5%;	Pred. No. 4.6e-35;		
Matches 209;	Conservative 115;	Mismatches 187;	Indels 197;	Gaps 27;

```

01 7 ELKQSSSRKKRSTASAPYDPBNDGIDPOSLEQGYTVQKKNRFTLSPISINHEKK 66
02 203 DUKLRSRARS--LASGWDDEDLDIDNDNDIPDYENKNGTIT---KDSIAVNR-EDSPAQ 255
03 67 GLTRYKSSPEKMWSTASDPEYFEKYTGRIKDNVSPEARHPLVAAPYIVHDEMITLSKN 126
04 256 GYKKLSIYSLESNTAGDPDYQKASGSPDKAIKAEARDPVLAAPYVGVGMEKLIISN 315
05 127 EDQSTQNDSEKRTISKTSRST--HISEVHGNAEVH---ANSTSTKTHSEVHGNAE 180
06 316 EHAST---DQKIVSRNTTNSKTSAMNAGVAINIAYONGFTSGTITNYSHTTE----- 365
07 181 VHAVIDHSLSLAGERTAETMGLTADTAPRANANIRVYNTGSTAPILYVLPPTSVLGKN 240
08 366 -NSTAVQMS---NGE-SMNTSLSIKSGESAYINAVNRYNTGTATAMKYTPPTINVL-DG 419
09 241 QTLATIKAKENQOILAPNNYPSKNIAPIALNADODESFTPTMNYNQELEKTYQOL 300
10 420 DTLITIKADQNOIGNLNPSNETEYPRKKGSLPLINTMDQFSPSLPIINYOQLKMLAGKIO 479
11 301 RLDPDQVYGNLATYFENGVRVDRD--GSWMSVLVPOIOETTRRIIFN-GKDLIVERRIA 358
12 480 KLEITVOVGN--YGIKKSOGOLITEGNSWSYISOLDSLSIILLDGSQ--VEERKAT 534
13 359 AVNPSDPLEETTKPDMTTLKEALKIAFGFNEPNQNOLOYOKDIYE--FDNFDOOTSQNTKN 416
14 535 AKDSSNPEDKTI-PVLTIGEAIEKARGATIKNGEILYFNGMPIDESCVELIFDGNFTANLIKE 593
15 417 QLAELNATINITYTVDKIKLNAKKNLIRD----- 445
16 594 RLNALNDKKIYVNV---OLERGMKILITRTSTYFNNFDDGYNPNPSSWSNVDSNNQDGLONA 649
17 446 -----KR 447
18 650 ANKLSGEIKIYIPMKLNPYKRYVSGYLKNSSTSNPTITVNIKAKEQKTYNLVSENDYKH 709
19 448 FHY-----DRNN-----IAYGADESUYVKEAH 468
20 710 FSYEPETIGRASNEIETLSGQIFLDNLSTTELNSTPEILKEPDIKVPDOEIT-DAH 768
21 469 REV-----INSTBEILLN-----IDKOIRKLISGVIYEI-DTEBLKEVINDRYDM 514
22 769 KKYVADISFNQSTANVYLDGLFEPTQTNKEVLADYQIKRYKATLEYSGGFQKIGTKRDEL 828

```



```
Db 326 ASANSHWOMSTYVD-----DTTGE-SFSQGSINTGESAYINPIRYN 370
OY 221 TGTARIYVNLPTSTSVLGNQTLATIKAKENQISOILAPNNYPSKNIAPLAIQODDS 280
Db 371 TGTAVYVNTPTTTIVIDK-QSVATIKGOESLIGDYLNCGYPTIIGEPMALNTMDOFS 429
OY 281 SPTITMNTNOFLEKTKQLRLDQOYGNATYVNFENGCRVFDGSMSEVLPQIOETT 340
Db 430 SRLIPINYNOLKSIDNGCVMLSTQSTGFNFAYN-SMGNLVTD-GNMNGPYLGTIKSTT 487
OY 341 ARII--FNGKDLNVERIAAVNPSDPLETTKPDMLKEALKIAPFNPNGNLQYOGKD 398
Db 488 ASLTSESGQTTQVA---VVAPESDPEDKT-PKLTLEQALVKAFALEKKNGKEYFHGLE 543
OY 399 IT---EPDFNPDQOTSONIKNOLAELMNTATYVLDKIKLNKNMILIRDKRPHYDRNI 455
Db 544 ISKNKEIQVFLDSNTNPNFENQLNKMTADKIMHCT-IKRN--NMILVAVITFKENISSI 599
OY 456 AV 457
Db 600 NI 601

RESULT 9
OY 09X377 PRELIMINARY: PRF: 204 AA.
AC 09X377:
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE PX01-111.
OS Bacillus anthracis.
OC Plasmid virulence plasmid PX01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=STERNE;
RA Okinaka R.T., Cloud K., Hamton O., Hoffmaster A., Hill K.K., Keim P.,
RA Koehler T., Lamke G., Kumano S., Mahillon J., Manter D., Martinez Y.,
RA Rieke D.O., Svensson R., Jackson P.J.;
RT "The sequence and organization of px01, the large Bacillus anthracis
RT plasmid harboring the anthrax toxin genes.";
RL J. Bacteriol. 0:0-0(1999).
DR EMBL; AF065404; AAD32415.1; -.
DR HSSP; P13423; IACC.
KM Plasmid.
SQ SEQUENCE 204 AA; 23029 MW; E1657B23AE4273FD CRC64;

Query Match 7.4%; Score 211.5; DB 2; Length 204;
Best Local Similarity 34.8%; Pred. No. 1.5e-05;
Matches 46; Conservative 31; Mismatches 34; Indels 21; Gaps 4;
OY 439 NMILIRDKRPHYDRNIIVAGADESVYKAEHREYINSSTEGLLNIDKDIRKLSGYIEI 498
Db 1 NMILVIRDE-YHDMNGNITVGDVSYLKNAAYKQILNMSSDGVSLNDEVDNALSGYMLQI 59
OY 499 EDTE-----GLKEVINDRYDMLNLSLRQDGKTFIDPKKRNKDKPLIYSNP 545
Db 60 KKPNSHLTNSPVTTTLAKRDSGVGELYRVL-----DOTGLDEKEDENMRSLV-DPG 112
OY 546 YKVNYYAVTKEN 557
Db 113 DDVYVYAVTKED 124

RESULT 10
OY 046149 PRELIMINARY: PRF: 2178 AA.
AC 046149; 046147; 046148;
```

```
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ALPHA-TOXIN.
OS Clostridium novyi.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC19402;
RX MEDLINE=95342160; PubMed=7616958;
RA Hofmann F., Herrmann A., Habermann E., von Eichel-Streiber C.;
RT "Sequencing and analysis of the gene encoding the alpha-toxin of
RT Clostridium novyi proves its homology to toxins A and B of Clostridium
RT difficile.";
RL Mol. Gen. Genet. 247:670-679(1995).
RN [2]
RP SEQUENCE OF 1204-2178 FROM N.A.
RC STRAIN=ATCC19402;
RA Hofmann F., Habermann E., von Eichel-Streiber C.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z48636; CA88565.1; -.
DR EMBL; Z23280; CA80818.1; -.
DR EMBL; Z23281; CA80819.1; -.
DR InterPro: IPR002479; CW_binding.
DR Pfam: PF01473; CW_binding_1; 10.
SQ SEQUENCE 2178 AA; 250134 MW; 9B0ADCE031C4A75A CRC64;

Query Match 6.3%; Score 181.5; DB 2; Length 2178;
Best Local Similarity 20.7%; Pred. No. 0.026;
Matches 122; Conservative 95; Mismatches 196; Indels 177; Gaps 27;
OY 63 HKKKGCTYKSSPEKMWASDPYDFEKYTRIDKNVSPEAR----HPLVAYPIVHWD 118
Db 320 NEKKLMNNY-----PYKMEQVPSIEKRIISFVNNDIINDILPLGDIKISQ 367
OY 119 ENILSKNEDSTONTSETRTISKNTS-TSRTHTSEVHG----NAEVHANTSTRTHT 172
Db 368 LEILSLRLKAATGKRTFSNAFIISNDSLTLLNLSIQLENREYELINSIOEKFKICEYD 427
OY 173 SEVHGNAEYHAVAIDHSLSLAGEKRTAETMG-LNTADARLKNANRY-----VMTGTAPI 226
Db 428 SYINSVSELVLETTPTKKNLSMDSSSEFYOOIGYLSGFPREVSVTFEGPNITSSATCDT 487
OY 227 VNVLPPT-SLVGKNQTLATIKAKENQISOILAPNNYPSKNIADIALNAODPFSTPI- 284
Db 488 YHFIKNTEDMLSSQDET-----FEASNNLYFSK-----THDEFKSSMLL 527
OY 285 --TMNYNOFLEKTKQLRLDQOYGNATYVNFENG-----VVDGSGSMSEVLPQIO 337
Db 528 RSNIAKEFEKQIKRITYIGR-----TLNYEDGLNFMNKKRYVT-----SELKAYIE 572
OY 338 ETTARIITNGKDLNVERIAAVNPSDPLETTKPDMLKEALKIFGENEPGNLQYOGK 397
Db 573 EYNSKRIYENDLNNI-----LQIOGDIDISYSAVNV-FGRN-PKKSILIOGV 618
OY 398 DITEEDPNPDQ--QTSQNIKNOLAELN-----AATNITYV 429
Db 619 DDFANVFFENGCIQVSDNINNLISRFNDIKIKTLILIGHENVEVPKLFGGKTVVDLTYN 678
OY 430 LDKIKLN-----AKNNIL-----TRDKRFH-----YD 451
Db 679 IKPKLQHLEREGVILKNKYKILNIILOGYMPFPRVDINSTPFVGLFNKISRDLPKGRS 738
OY 452 RNNIIVAGADESVY---KEAHREYINSSTEGLLNIDKDIRKLSGYIV---EIDTEGK 505
Db 739 KNQLEISANKYAIRINREKREVLDYFGK-WVSNNDLIAEQISNKRYVYVWMEVENT---- 793
OY 506 EYINDRYDMLNLSLRQDGKTFIDPKKRNKDKPLIYSNPYK---VNYVA 552
Db 794 --LSARVEQLN-----KVAEFAKIDINSITQTTNDELKOSLVNYTA 832
```

```
RESULT 11
ID 026021 PRELIMINARY; PRT; 604 AA.
AC 026021:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE STAP ANTIGEN.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=79/96;
RX MEDLINE=95021499; Pubmed=7935600;
RA Fiodok D.A., Bottius E., Braham K., Moelans I.I.M.D., Alkawa M.,
RA Konings R.N.H., Cerna U., Olafsson P., Kaldoh T., Asavanich A.,
RA Guerlin-Marchand C., Druilhe P.;
RT Cloning and characterization of a novel Plasmodium falciparum
RT sporozoite surface antigen, STAP.
RL Mol. Biochem. Parasitol. 64:219-232(1994).
DR EMBL: Z26314; CAA81224.1;
SQ SEQUENCE 604 AA; 67051 MW; DE23D9442A509667 CRC64;

Query Match 5.8%; Score 166; DB 5; Length 604;
Best Local Similarity 21.3%; Pred. No. 0.037;
Matches 125; Conservative 83; Mismatches 250; Indels 128; Gaps 28;

QY 45 TYDVAKKRRFLSPWISN-----IHEKKG---LTKKSSPEKKSTASD--PYSD 87
DB 39 TYVKKHNR-FLSEYQSNFLGGYSAALKLVNSKKSGTNVTKYNS--ENTNTNNIPSS 95
QY 88 FEKVTRIDRNVSPKARHPVLAAYPIVHVDENILIKSNED-----OSTONTSEPTIS 142
DB 96 SYTYNRRLAANS-----TTSTTKYTDNKKTKIKLGNSTIINTSTENT--SATKVT 149
QY 143 K-----NTSRTHTSEVHGNAEYHANTSTSRHTSEVHGNAEYHAVAIDHSL 191
DB 150 ENVITNQLITGNNTTNTSTTEHNNINNTNSTENTSTATKKEVNVITNQLITGNNT 209
QY 192 LAGERTWAEHTMGINT---ADTARLANINRYVMGTAP--IYVLPPTSLVLGKQTLAT 245
DB 210 TTNFTSTTEHNNINNTNSTNSNTNTNLT-DNTSTTKKLTDNINTQNLTTSTNTTYS 268
QY 246 -----IKAKENQLSQILAPNNY-----YPSKNLAPIALNADDESPPTMYNV--QF 291
DB 269 TDNNNTINTKPIDNNNTDIKFTDYNITGKTFDKKN---TDIKADNNNTITTTDNTNTY 325
QY 292 LELEKTKQLRLTDQVYGN-IATYFNENGVRVDTGSGNMSEVLPOIETTARILFNGKDL 350
DB 326 ISTDNSKKTNYISTDNSKKTNYISTDNDAOTILTDNDNNTDILILD-----NNNTDT 376
QY 351 NLVERKIAAVNPSPDELTTPDMTKLEALKIAGFNPKNGLQYOGKDITEFDNFPQOT 410
DB 377 ISTDNDAOTKATDNNNTTKATDNNNTKI---SPDNN---NPKTSTDDNNNTNT 428
QY 411 SQRIKQALALNATNTYTVDKIKLAKKMMILIRDKRFPHDRNNI--VVGADSESVKKAHR 469
DB 429 KATDNNNTKTIISTDNNNT-----KTIISTDNNNTKTIISTDNNNTKTIISN 471
QY 470 EVINSST-----EGLLN-----IDKDIRKILSGYIEIDTEGLKEVIN--- 509
DB 472 DNNNTNTISTDNNNNMTNQVFAANNYNETSDELDKDCDY---SEEKEIKSMINAYL 528
QY 510 DRKDMNLISLRQDGKTFIDFKKYNDKPLXYISPNKVVAVYATK 555
DB 529 DKLDLETVKRIHSDISTGLE-KKNRP--NOIITHLNLNKKMYNLIK 571

RESULT 12
```

```
025579
ID 025579 PRELIMINARY; PRT; 2529 AA.
AC 025579;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE TOXIN-LIKE OUTER MEMBRANE PROTEIN.
GN HP0922.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group.
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; Pubmed=9252185;
RA Tomb J.F., White O., Kerlavage A.R., Clayton R.A., Sutton G.C.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-B., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weldman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
DR EMBL: AE000602; AAD07969.1;
DR TIGR: HP0922;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 2529 AA; 274561 MW; 440882E8644472EC CRC64;

Query Match 5.7%; Score 164.5; DB 2; Length 2529;
Best Local Similarity 19.4%; Pred. No. 0.033;
Matches 113; Conservative 83; Mismatches 206; Indels 181; Gaps 26;

QY 120 NIIISK--EDOSTONTSETRITSKNTSTSRHTSEVHGNAEYHANTSTSRHTSEVHG 177
DB 825 NIYLTNFKTGEVSNBDDGGANITPKASDNITMDGLNYNDK-----TVTKMIQT 875
QY 178 NAEVHAVAIDHSL-----SLAGERT---AETMGLTFAD-----TAR 211
DB 876 GASQSHATPDALNNISYTNSSFSQMTWGRFSKAKINFSNAFSGFTNPGSSVISA 935
QY 212 LMANIRYVNT--GTAPIYVNLPTSLYLGKQTLATI-----KAKENQLSQILAPNNYYP 264
DB 936 ATNLSFINSRLNGCAYVN-LQANSLIFNNTOAVFNVLSRGTSPNATQQLGNTFTL 994
QY 265 S-----KNLAPIAL--NAODDESPPTMYNVNOFLELEKTKQLRLTDQVYG 309
DB 995 SSQSLNPNFGDTTLQNNANNTILGNKSOAAFKNS--LTDNNSNLSLDQSVNANNTSAFN 1053
QY 310 NIATYFNENGVRVDTGSGNMSEVLPOIETTARILFNGKDLNLT--VERRIAVNPSPDET 368
DB 1054 NQASLNINYS-----QATFNSLFFNGKTLISLWASSKSLWASNSPNSNT 1097
QY 369 TKPDMTKLEALKIAGFNEPNGLQYOGKDITEF-----DFNPDQOSTONIKQOLA 419
DB 1098 T---INLDDSVLSASNTSSLNANINFGASQADGCGNTIIDTASPNFDSASLNFNMLTA 1154
QY 420 -----ELNATNTYTVDKIKLAKKMMILIRDKRF- 448
DB 1155 NGALNPNGYPSLTALKALMSVSGQFVLGNNGDINLSDI--NIPDNITKSTYVILNAOKGIT 1213
QY 449 -----HYDRNNIAYGA-----DESVKKAHR-----EYVN- 473
DB 1214 GISGANGYEKLFYGMKIQNATYSDNNNIQTWSFINPLNSQIIQIESIKNKDILTEVLNN 1273
QY 474 -SSTEGLLNIDKDI-----RKILSGYIEIDTEGLKEVINRYDMNLISL- 520
DB 1274 PMSASNTIFNPAPLRYNQASKNPPTGYSDYSDNQK-----GTYVYLSNINIKGLFTPKGS 1328
```

QY 521 --RODGTFIDFKKYNKLPY---ISNPNKYVNYAVTKENT 558
DB 1329 QTPQAPGTSPFNOPPLSSILNTYKNGFSSENLK-TLLGILSONS 1370

RESULT 13

09P008 ID 09P008 PRELIMINARY; PRT; 4688 AA.
AC 09P008;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHEICAL PROTEIN U0482.
GN U0482.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROVAR 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
urealyticum.";
RL Nature 407:757-762(2000).
DR EMBL: AE002145; AAF30894.1; "
DR InterPro: IPR001152; Thymosin_b4.
DR InterPro: IPR000626; Ubiquitin.
DR Pfam: PF00240; ubiquitin; 1.
DR SMART: SM00152; THY; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 4688 AA; 534880 MW; B53ABFAFEE1997E CRC64;

Query Match 5.58; Score 158; DB 2; Length 4688;
Best Local Similarity 21.2%; Pred. No. 1.9;
Matches 138; Conservative 107; Mismatches 270; Indels 136; Gaps 32;

QY 1 DNLOJPELKOKSSN-----SRKRSFSAQPIVPPDR-----NDGIIPD 37
DB 3803 DNLM-PEKRYKLENTSELSPLKTHNLVSLIDKENISLITETGPAVLKIQTQNDTIND 3861
QY 38 SLEVEGYTVADYKRRTPSPWISNIHEKKGLTKYKSSPEKMTASDPYSEKVGRIK 97
DB 3862 TQGTINVLGSGNSK-YNGRQIKVYKDNNNVITYESS---LITLQKNDIYOLLNSLNS 3917
QY 98 NVSPARHPPLVAAPYIVV-----DME-----NIIISKNEOSTONTDSE-----TR 139
DB 3918 N-----REYREKEIENIHISNTNNEFELEKNGVSNFTQTKNTYVOMNDSATIVGTR 3972
QY 140 TISKNTSRTHTSEVHGNAEY---HANTSTRTHTSEVHGNAEYHAAVIAIDHSLAGER 196
DB 3973 GVNREFKI-KSEDKLENNQOVANFAFKETIKROTNTMLOTRIPLKDYTSDEK-----EG 4026
QY 197 TWAETMGLNT---ADTARLANIRVYNTGTAPYVNLPTTSLVLKQNTLA-----TIK 247
DB 4027 TWAHDLNSVNFKEETTKYKLVKIOFVNKPRTAKNNINNSENNVLIDNINSINSNVEFTK 4086
QY 248 AKENOLSLAPNNYPSKKNLAPALN-AODEFSPTITMAY--NQFLELEKTKQLRLDT 304
DB 4087 VGDHKLIMITSSNNVNTSQTINFTLSGVKSKSWGKKIKLSKSDTSESIHTNEVLIES 4146
QY 305 DOYVGNINTYPENGR-----VRVDTSNNSEVLPOIQTETARILIFNGKDLNVERRIA 358
DB 4147 NKTQYNIILNKKRNYITLIDVKLIDNNNSDPEKESNLSTFTTTSALINVEIEI 4206
QY 359 AVNPSDPLETT-----KPDWTLKEA-----LKIAFGFNPNNGNLQYOGDITEF 402
DB 4207 SNRSTNLKSTIIKINLNDPNNVLRDKDOATIVYGNKKQAMGFIVSGNIKILVTATLVDL 4266

QY 403 DPNFDQTSQNIK-NQALNATNTIYVLDKIKLAKANNILIROKRFHYD-----RNN 454
DB 4267 NFN-DKVINIVISFNKKPSIAEN-----IGDKSNII-----YNNDSIPKLEIND 4313
QY 455 IAYGA---DESVYKKAHEVINSSEGLLNDIRKILSGYIETDGLKEVIND- 510
DB 4314 IIVNGPIKKEIYKNAK--NNIDVDLGLQINPKIAHLNR-FLAKFKSTN-----NDI 4364
QY 511-RYDMLNITSL-RODGTFIDFKKYNKLPYISPNYK-VNVAVTKENT 558
DB 4365 IETVINGSSLVNNDGKTSIRFTLNLK-----ANKXLSVDVYLVVNNNS 4410

RESULT 14

096129 ID 096129 PRELIMINARY; PRT; 1308 AA.
AC 096129;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PREDICTED MEMBRANE ASSOCIATED PROTEIN.
GN PFB0125C.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=3633;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tettein H., Carnuci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shallem S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.";
RL Science 282:1126-1132(1998).
DR EMBL: AE001374; AAC11815.1; "
DR InterPro: IPR001313; PDM.
SQ SEQUENCE 1308 AA; 155585 MW; 9722F0336606C366 CRC64;

Query Match 5.58; Score 157; DB 5; Length 1308;
Best Local Similarity 19.9%; Pred. No. 0.37;
Matches 130; Conservative 95; Mismatches 219; Indels 210; Gaps 31;

QY 47 DVKNKRTFLSPWISNIHEKKGLTKYKSSPEKMTASDPY----- 85
DB 42 DVEKKDVTL-----NLDKKNVEEYKKNKDVFKNEDEFFVFDDKEINKLKIKEGON 96
QY 86---SDFEYVGRV--DKNVSPARHPPLVAAPYIVHDMENIILSKNEOSTONTDSETRT 140
DB 97 MKENEFINEKEGYILNDEVNS-TINNITSLNNDILHSSDKNCTSYNIPSGNNNNN-- 153
QY 141 ISKNSTSTRTHTSEVHGNAEYHANTSTRTHTSEVHGNAEYHAAVIAIDHSLAGE----- 195
DB 154 -----NNNNNVHSNNS-----NIFVDSHMEHFDDITDTEFEKID 187
QY 196 RTWAETMGLNTADTARLANIRVYNTGTAPYVNLPTTSLVLKQNTLATIKAKENOLSQ 255
DB 188 QTNFSEFPQNTSEFKKKVNEEELMKHTDNT-NI---CDKIIDK-----KKKNNTLSID 236
QY 256 ILAPNNYPSKKNLAPALNAODEFSPTITM--NYNOFLELEKTK-OLRLDTDOYVGN 311
DB 237 LIHDNLFDNMLNITYEDNNKNDVISTDLMKKNYK--NFEKNEIDVYVDTSTTFENI 294
QY 312 ATYENENRVRVDGSSNSEVLPOIQTET-----ARIIFNGDLNVERRIA- 359
DB 295 NNDNNEKNLMLNMQSDKEELNANNKDDTFYINNKFLISENNIILEDKDISFIDKIESN 354
QY 360 -----VNPSDPLETT---PDMWTLKE-----AL 379
DB 355 KCEDYCVNNNNNNNNRNLSDILENAYSKDCESRTINEDRIYNNFEDMDKISHDAFDIT 414

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:52:29 ; Search time 44.8 Seconds
(Without alignments) 456.674 Million cell updates/sec

Title: US-09-747-521-4_COPY_178_735
Perfect score: 2871
Sequence: 1 DNLQPELKOKSSNRKRRS.....LYISNPYKVNYAVTKENT 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	2758	96.1	764	1	PAG_BACAN
2	211.5	7.4	192	1	YPA_BACAN
3	150	5.2	1230	1	SMC3_YEAST
4	141	4.9	1658	1	YM67_YEAST
5	139	4.8	1358	1	SIR4_YEAST
6	138.5	4.8	1251	1	RBP2_PLAIVB
7	138.5	4.8	1928	1	MYSL_YEAST
8	138	4.8	1385	1	FAT1_SCHPO
9	136	4.7	1790	1	USO1_YEAST
10	136	4.7	1803	1	YJL3_YEAST
11	134	4.7	753	1	YBV2_YEAST
12	134	4.7	1570	1	P3K1_DICDI
13	132.5	4.6	678	1	SIR1_YEAST
14	132	4.6	1208	1	PCP1_SCHPO
15	130.5	4.5	633	1	IPAA_SHIFL
16	130.5	4.5	803	1	SMIG_YEAST
17	130	4.5	2334	1	MAPA_BACSU
18	130	4.5	2710	1	TOXA_CLODI
19	129.5	4.5	719	1	YMA1_YEAST
20	129.5	4.5	2869	1	RBP1_PLAIVB
21	129	4.5	1182	1	CGA2_HELPY
22	128.5	4.5	1271	1	Y338_MYGE
23	128.5	4.5	1460	1	N159_YEAST
24	128	4.5	976	1	SCP1_HUMAN
25	128	4.5	1167	1	CAGA_HELPY
26	127	4.4	1744	1	TANA_XENTIA
27	126	4.4	1487	1	MDS3_YEAST
28	125.5	4.4	1630	1	MSPI_PLAFK
29	125.5	4.4	1639	1	MSPI_PLAFW
30	125	4.4	1726	1	MSPI_PLAFP
31	124.5	4.3	635	1	PIB2_YEAST
32	124.5	4.3	678	1	YNC7_YEAST
33	124	4.3	597	1	PTPX_CANAL

ALIGNMENTS

RESULT	1	STANDARD	PRT	764 AA.
PAG_BACAN				
ID	P13423			
AC	01-JAN-1990 (Rel. 13, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	PROTECTIVE ANTIGEN PRECURSOR (PA) [CONTAINS: PA-20; PA-63].			
PAG				
OS	Bacillus anthracis.			
OG	Plasmid pX01.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
ON	Bacillus/staphylococcus group; Bacillus.			
OX	NCBI_Taxid=1392;			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91172073; PubMed=3148491;			
RA	Welkos S.L., Lowe J.R., Eden-McCutchan F., Vodkin M., Leppla S.H.,			
RA	Schmidt J.J.;			
RT	"Sequence and analysis of the DNA encoding protective antigen of			
RT	Bacillus anthracis.";			
RL	Gene 69:287-300(1988).			
RN	[2]			
RP	DOMAINS.			
RX	MEDLINE=91332080; PubMed=1651334;			
RA	Singh Y., Kimpel K.R., Quinn C.P., Chaudhary V.R., Leppla S.H.;			
RT	"The carboxyl-terminal end of protective antigen is required for			
RT	receptor binding and anthrax toxin activity.";			
RL	J. Biol. Chem. 266:15493-15497(1991).			
RN	[3]			
RP	X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).			
RX	MEDLINE=97192099; PubMed=9039918;			
RA	Petosa C., Collier R.J., Kimpel K.R., Leppla S.H., Liddington R.C.;			
RL	"Crystal structure of the anthrax toxin protective antigen.";			
RL	Nature 385:833-838(1997).			
CC	- FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,			
CC	AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE			
CC	DEATH. PA IS THOUGHT TO BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC			
CC	CELLS, THEREBY FACILITATING THE INTERNALIZATION OF LF OR EF. PA			
CC	ASSOCIATED WITH LF CAUSES DEATH WHEN INJECTED. PA ASSOCIATED WITH			
CC	EF PRODUCES EDEMA. PA INDUCES IMMUNITY TO INJECTION WITH ANTHRAX.			
CC	- SUBUNIT: ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT PROTEINS, A			
CC	PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN EDEMA FACTOR			
CC	(EF). NONE OF THESE IS TOXIC BY ITSELF. ONCE ACTIVATED, PA FORMS			
CC	HEPTAMERS WHICH INSERT INTO MEMBRANES AND FORM CATION-SELECTIVE			
CC	CHANNELS.			
CC	- SUBCELLULAR LOCATION: SECRETED.			
CC	- DOMAIN: THE C-TERMINAL PART OF PA IS REQUIRED FOR RECEPTOR BINDING			
CC	AND TOXIC ACTIVITY.			
CC	- PTM: PROTEOLYTIC ACTIVATION BY URIN CLEAVES THE PROTEIN INTO TWO			
CC	PARTS, PA-20 AND PA-63, THE LATER HEPTAMERIZE.			
CC	- SIMILARITY: TO C. PERRINGENS TOTA-B TOXIN AND TO VIP1 TOXINS IN			
CC	BACILLUS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: M22589; AAA22637.1; .
DR PDB: IACC; 11-FEB-98.
DR InterPro: IPR003896; Binary_tox.
KW Toxin; Plasmid; Calcium-binding; Signal; 3D-structure.
FT SIGNAL 1 29
FT CHAIN 30 764 PROTECTIVE ANTIGEN.
FT CHAIN 30 196 PA-20.
FT CHAIN 197 764 PA-63.
FT DOMAIN 30 287 DOMAIN 1, CALCIUM-BINDING.
FT DOMAIN 288 516 DOMAIN 2, HEPTAMERIZATION.
FT DOMAIN 517 764 DOMAIN 3.
FT CA_BIND 206 206
FT CA_BIND 208 208
FT CA_BIND 210 210
FT CA_BIND 217 217
SQ SEQUENCE 764 AA; 85811 MW; 84BB22690FEAAB5 CRC64;

Query Match 96.1%; Score 2758; DB 1; Length 764;
Best Local Similarity 96.8%; Pred. No. 1,8e-142;
Matches 540; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

OY 1 DNLQPELKQKSSNRKRRKSTASGPTVPDRNDGIPDSLEVGYYVDVKNKRRFLSPWIS 60
DB 178 DNLQPELKQKSSNRKRRKSTASGPTVPDRNDGIPDSLEVGYYVDVKNKRRFLSPWIS 237
OY 61 NTHKKKGLTKYSSPEKNTASDPYSDEPKVGRIDKNVSPARHPVAAYPIVHVDEN 120
DB 238 NTHKKKGLTKYSSPEKNTASDPYSDEPKVGRIDKNVSPARHPVAAYPIVHVDEN 297
OY 121 ILSKNEQOSTONTSETPTTISKNTSTSRHTSEVHGADEVANSTSTHTSEVHGADE 180
DB 298 ILSKNEQOSTONTSETPTTISKNTSTSRHTSEVHGADEVANSTSTHTSEVHGADE 357
OY 181 VHAVALDHSLSLAGERTVAETMGLTADTARLANIRVYNTGTAPLYNVLPPTSLVLGKN 240
DB 358 SSTVAIDHSLSLAGERTVAETMGLTADTARLANIRVYNTGTAPLYNVLPPTSLVLGKN 417
OY 241 QTLATITAKENQSLQILAPNNYPPSKNLAPILANODDSSPTITMNYNOFLEKTKOL 300
DB 418 QTLATITAKENQSLQILAPNNYPPSKNLAPILANODDSSPTITMNYNOFLEKTKOL 477
OY 301 RLDTQVYGNATYVNFENGVRVDTGSNMSEVLPQIQETTRARIIFNGKDLNVERRIAAV 360
DB 478 RLDTQVYGNATYVNFENGVRVDTGSNMSEVLPQIQETTRARIIFNGKDLNVERRIAAV 537
OY 361 NPSDPLETTKPDMTLEKALIAFGFNEPENGNIQYOGKDTTEFDNFNDQTSQNIKNQALAE 420
DB 538 NPSDPLETTKPDMTLEKALIAFGFNEPENGNIQYOGKDTTEFDNFNDQTSQNIKNQALAE 597
OY 421 LNATNIVYVLKIKIRKNAKNILIRDKRFHYDRNNIIVAGADESVVAEAREVINSSTEGIL 480
DB 598 LNATNIVYVLKIKIRKNAKNILIRDKRFHYDRNNIIVAGADESVVAEAREVINSSTEGIL 657
OY 481 LNIDDKIRKILSGYIYEIEDTGLKEVINDRYDMINISLRDQGTGFIDEFKYNNKPLPY 540
DB 658 LNIDDKIRKILSGYIYEIEDTGLKEVINDRYDMINISLRDQGTGFIDEFKYNNKPLPY 717
OY 541 ISNPNYKVNYAVTKENT 558
DB 718 ISNPNYKVNYAVTKENT 735

RESULT 2
YPA_BACAN STANDARD: PRT; 192 AA.
AC P13422;

DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE HYPOTHETICAL 21.6 KDA PROTEIN IN PROTECTIVE ANTIGEN 5' REGION.
OS Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89172073; PubMed=3148491;
RA Welkos S.L., Lowe J.R., Eden-McCutchan F., Vodkin M., Leppla S.H.,
RT Schmidt J.J.
RT "Sequence and analysis of the DNA encoding protective antigen of
RT Bacillus anthracis."
RL Gene 69:287-300(1988).
CC -I- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: M22589; AAA22636.1; .
DR HSSP: P13423; IACC.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 162 182 POTENTIAL.
SQ SEQUENCE 192 AA; 21620 MW; DBC7150AE78F8AFA CRC64;

Query Match 7.4%; Score 211.5; DB 1; Length 192;
Best Local Similarity 34.8%; Pred. No. 2.7e-05;
Matches 46; Conservative 31; Mismatches 34; Indels 21; Gaps 4;

OY 433 NMLIRDKRFHYDRNNIIVAGADESVVAEAREVINSSTEGILLNIDDKIRKILSGYIYEI 498
DB 1 NMLIRDKRFHYDRNNIIVAGADESVVAEAREVINSSTEGILLNIDDKIRKILSGYIYEI 59
OY 499 EDTE-----GLKEVINDRYDMINISLRDQGTGFIDEFKYNNKPLPYISPN 545
DB 60 KPSNHLNPSPTITLAGKDSGVGELYRLS-----DGAGFLDKPKDENKRSIV-DG 112
OY 546 YKVNYYAVTKEN 557
DB 113 DDVYVAVTKED 124

RESULT 3
SMC3_YEAST STANDARD: PRT; 1230 AA.
AC P47037;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CHROMOSOME SEGREGATION PROTEIN SMC3 (DA-BOX PROTEIN SMC3).
GN SMC3 OR YJ1074C OR J1049.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W303;
RX MEDLINE=97474309; PubMed=9335333;
RA Michaelis C., Ciosk R., Nasmyth K.;
RT "Cohesins: chromosomal proteins that prevent premature separation of
RT sister chromatids."
RL Cell 91:35-45(1997).
RN [2]
RP SEQUENCE FROM N.A.

RA Rose M., Koetter P., Enlian K.D.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RA Sor F.J.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
 CC PART OF A CHROMOSOME CONDENSATION MOTOR.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
 CC FLANKED BY PUTATIVE GLOBAL REGIONS AT THE N- AND C-TERMINUS.
 CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y14278; CAA74655.1; -
 DR EMBL: 249349; CAA89366.1; -
 DR EMBL: X88851; CAA61313.1; -
 DR SGD: S0003610; SMC3.
 DR InterPro: IPR003459; ABC_transport.
 DR InterPro: IPR003405; SMC_C.
 DR InterPro: IPR003395; SMC_N.
 DR Pfam: PF02483; SMC_C; 1.
 DR Pfam: PF02463; SMC_N; 1.
 DR Mitosis: ATP-binding; Coiled coil; Nuclear protein.
 FT NP_BIND 32 39 ATP (POTENTIAL).
 FT DOMAIN 172 482 COILED COIL (POTENTIAL).
 FT DOMAIN 685 1041 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1230 AA; 14136 MW; B152D88F780341F CRC64;

Query Match 5.2%; Score 150; DB 1; Length 1230;
 Best Local Similarity 20.4%; Pred. No. 0.73;
 Matches 18; Conservative 102; Mismatches 255; Indels 182; Gaps 28;

QY 3 LQPELKKOKSSNRKSTASGPTVPDRNDGIPDSLEVEGYTVYVKNKRTPLSPWISNI 62
 DB 371 LQPLASLQOKRDLILKGEYARKSKDERDTWHSEIT-----ELKSS-----IQNL 417
 QY 63 HE-KKGLTKYKSSPEKWSSTASDPYSDEKYTGRIKKNVSEARHPLVAAYPIVHDMENI 121
 DB 418 NELESQLOMDRTSLRKQYSID-----EIELELIDISNGPDTGQL-----EDFDESLI 466
 QY 122 ILSKNEDQSTQNTDSETRTSKNTSTSRHTSEYHGNAEYHANTSTSRHTSEYHGNAEV 181
 DB 467 HLOKQKLESELDTRKELMRKQKLOTLETLLSPVNON-QRNVNEMTMSRS-----LANGII 520
 QY 182 HAVAIDHSLSLAGE-----RTMAETMG-----LNTADTARLANNIR 217
 DB 521 NKKEITTEKLTSPESVFGTIGELIKYNDKRTCAEVLGGNSLFIHTVETETATLNNEL 580
 QY 218 YVNTGTAPIVNLPPTSLVLGKNOTLATIKAKENOLSQLAPNNYPSKMLAPITALAOD 277
 DB 581 YRMKGGGVNPE--IPLNKLSLSDSVKFPSSNTTQIQFPLPKIKIYERFKA-----VKH 633
 QY 278 DESSTPTTMYNOFLEKTKQLR--LDTDQV-YGNIAF--YVFENGRAVDTGSMSE 331
 DB 634 VFEKTIYVKNLGGOLKAKKHKLNATLTDGDRADKRGVLGGLDQHKRTRLESILKMLNE 693
 QY 332 VLPOIOETARTIFNGCLDLVERRIAANPS-----DPLETTKPD 372
 DB 694 SRQOHKILIEELDFVRNELNDIDIKIQVNGNIRKKSNDRESVLTINLEYRTSLNTKKE 753
 QY 373 -MTLKEALK-IAFGFNPNGLLOYOGKDIYEFEN--EDQOTSNIKOLA----- 419
 DB 754 KLIEESLNLMIILIKLEKLTNNRTFAQEKLTMTFENDLLQDFDSLSKKEKELSELTKEIS 813

QY 420 -----ELNATN-----ITYVLDKIKLAKKNNILIRDKRFHYDRNNIANG----- 458
 DB 814 AAHNKLNITSDALEGITTTID--SLNAELSKLIPENDLESKMSFVGADPFIQLODELK 871
 QY 459 -----ADESVYKAEHREVINSST-----EGLL-----NIDK 485
 DB 872 ELQLEKESYKQEHENAVLEIGTYOREIESLIAETNNKKLLEKANNQORLLIKLNDFOK 931
 QY 486 DIRKILSGYIEEDTEGLE-----VINDRYDM-----IN-----ISL 520
 DB 932 SVEKTMIKTKTLTVTRRELQORIREIGLPEDALVNDPFSITSDQLQRLNDMTEISGL 991
 QY 521 RQDCK-TEIDPKKYNDK 536
 DB 992 KNVKKRAFNFKFKNER 1008

RESULT 4
 ID YM67_YEAST STANDARD; PRT; 1658 AA.
 AC Q03661; Q04988;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOThETICAL 187.1 KDA PROTEIN IN GUA1-ERG8 INTERGENIC REGION.
 GN YMR219W OR YMR261.13 OR YMR959.01.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID:4932;
 RN [1]
 RP SEQUENCE OF 1-711 FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
 RA Walsh S.V.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 608-1648 FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 249809; CAA89934.1; -
 DR EMBL: 249939; CAA90190.1; -
 DR SGD: S0004832; YMR219W.
 KW Hypothetical protein.
 SQ SEQUENCE 1658 AA; 187137 MW; 3893P968305A57D CRC64;

Query Match 4.9%; Score 141; DB 1; Length 1658;
 Best Local Similarity 20.4%; Pred. No. 3.4;
 Matches 125; Conservative 92; Mismatches 221; Indels 176; Gaps 31;

QY 11 KSSNSRRKSTASGPTVPDRNDGIPDSLEVEGYTVYVKNKRTPLSPWISNIH-E-K 65
 DB 488 KSEYSEVWLENEDTPAIVERENC-IND--VEGYDVYKSVESDLHESP--DWLYDLAA 541
 QY 66 KGLTKYKSS-----PEKWSASDPY--SDEKYTGR-IDKNVSEARHPLVAAYPIVHV 116
 DB 542 RAMLQOQSRNSNCPQKEBOVSEYLGHSNGSNLSGSLDES---EEOIPLKD----- 591
 QY 117 DMENIILSKNEDQSTQNTDSETRTSKNTSTSRHTSEYHGNAEYHANTSTSRHTSEV 175
 DB 592 -----FTGNNNNNLATDRGDLSSVIEIEYKSEKKLDOSTFKELVLPSTDTINNSS 644

```

OY 176 HGMAEYHVAIDHSLAGERTWMTMGLNTRA-----DTARLNANI 216
DB 645 LGNESTIYSLDDADAISENLDPVLEIKTPPKYEVISSVYSTSEDTVAMPPOV 704
OY 217 RYVNGTAPIYVLPPTSLVLGKNOTLATIKAKENQLSQILAPNNYPSKNIAPIALNAQ 276
DB 705 EY-----TSPFNMD -PFNSL-----ND---DYKKHLLKSTLAA-----LAPAFYTKD 744
OY 277 DDFSSPTITM-----NYNOFLEKTKOLRLDTPDOVYGNIAFYNGFNRVAVDTGSN 328
DB 745 AEFVAGVYTKSCLTSTSGHTNIFHTSKETKQVS -DLDESTENVTPEENWTG---DENKN 799
OY 329 MSEVLPOQETARTIIFNGKDLNVERIAVNPSPDLETTRKPDITLKEALKIAGFNEP 388
DB 800 QSKNFPVGNASTDKSTENTND---EKYFSAINNTN---VTGSSCEDDIETASNAVEE- 850
OY 389 NGNLQYQKGDITEFDFNF -DOOTSNN---IKNOLA-----ELNATNIYVLDIKIKNA 437
DB 851 --NLATYCKDMNHEAMSSGDECVKQNDGSKTQISFSTDSPDNPQESNDNTEFSSTK--- 905
OY 438 KMNILIRDKRPHYDRNNIAVGADESVEKRAHR-EVIN-----SST 476
DB 906 -----YKVRNSDLEDESLKKEIKAEVYDKDEESEDSEYEQDYADPEPCND 953
OY 477 EGLLNLIDKDIRKILSGYIVEIEDEGELKEVINDRYDMLNISLQDQKTFIDPKKYNDK 536
DB 954 EGSNENIYKGTJK-----DTLGIVEPENEKVN-----KVHEE 985
OY 537 LPLYISNPYKYNV 550
DB 986 ETLFEANVSSSVNV 999

RESULT 5
SIR4_YEAST
ID SIR4_YEAST STANDARD: PRT: 1358 AA.
AC P11978:
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE REGULATORY PROTEIN SIR4 (SILENT INFORMATION REGULATOR 4).
GN SIR4 OR STE9 OR ASD1 OR UTH2 OR YDR227W OR YD9934.12.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
ON NCBI_TaxID=4932;
OX (1)
RX SEQUENCE FROM N.A.
RP MEDLINE=88142836; PubMed=3325825;
RA Marshall M., Mahoney D., Rose A., Hicks J.B., Broach J.R.;
RT "Functional domains of SIR4, a gene required for position effect
RL regulation in Saccharomyces cerevisiae.";
RL Mol. Cell. Biol. 7:4441-4452(1987).
RN (12)
RP SEQUENCE FROM N.A.
RX MEDLINE=95192063; PubMed=7885847;
RA Davies C.J., Hutchinson C.A. III;
RT "Insertion site specificity of the transposon Tn3.";
RL Nucleic Acids Res. 23:507-514(1995).
RN (13)
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE PROTEINS SIR1 THROUGH SIR4 CONTROL THE EXPRESSION OF
CC MATING TYPE GENES RESIDENT AT LOCI OF EITHER END OF CHROMOSOME
CC III. SIR3 AND SIR4 ASSOCIATE WITH THE C-TERMINUS OF RAP1 TO FORM A
CC DNA-BINDING COMPLEX THAT INITIATES THE REPRESSION AT THE HM LOCI
CC AND TELOMERES.
CC -1- SUBUNIT: INTERACTS WITH RAP1 C-TERMINUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DB 431 -----NYNSTSRST-----EKKNDNTSAKKN--NGENKKKGKRPPELMSSTEAHAKVTE 478
OY 475 STEGLNLINDIRKILSGYIVEIEDEGELKEVINDRYDMLNI--SSLRQDGKTFIDPKK 532
DB 479 ETTKOIQSVRLDGRKVL-----QKVG-ESHIDSRNNTLNVPSKRPQGLGELPNPKK 530
OY 533 Y--NDKLPYISNPYKYNVAVYATKE 556
DB 531 HKPNEGRTPNISNGTINIQKLEPRE 556

RESULT 6
RBP2_PLAVB
ID RBP2_PLAVB STANDARD: PRT: 1251 AA.
AC 000799:
DT 01-APR-1993 (Rel. 25, Created)

```

```

Query Match 4.8%; Score 139; DB 1; Length 1358;
Best Local Similarity 19.3%; Pred. No. 3.3;
Matches 121; Conservative 107; Mismatches 234; Indels 164; Gaps 24;
OY 2 NLQDPELQKSS-----NSRKKRSTSGPTVPDRDNDGIPDSLEVEGYVD 47
DB 24 NDKITREKESNEVTKPIPLTFPAKSKNYGRSPAIHTSHQSDVAPTS----- 74
OY 48 VKNKRTPLSPIWISNHEKKGTLKYKSSP-EKWSPTASDPYSDFEKT----- 92
DB 75 -----HKOLOQPKSSPLKNNVNSPFSNLEKISNKLKLSLRKTS 117
OY 93 GRID-KNVSPEARHPVLAAYPIVAVDMENIILSKNEDSTONTQSE-TRITSKNTSRT 150
DB 118 GRIESNPSHDSASRLAS-----FEOTAFSRRAOQOSTFNSKPVRTIVP-ISTSYQT 168
OY 151 HTSEVHGNAEYHVAANTSTRTHTSEVHGNAEYHVAIDHSLAGERTWMTMGLNATADTA 210
DB 169 NNSFLSGVKSLLSPEKIRDSKEILG-----INLANQPLVE----- 205
OY 211 RLMAIRYVNTGTAPIYVLPPTSLVLGKNOTLATIKAKENQLSQILAPNNYPSKNIAP 270
DB 206 -----KPLKKGSAID-----GASVSLTKDKSIRKDYEEKKEKLNKFNASDSLSV 255
OY 271 IALNAQDFFSPTITM-----NYNOFLEKTKOLRLDTPDOV 308
DB 256 PKVSGADSGISPEESKARSPGIARPNALQTEVYGINEESTNERLEINQEKPVKLDENSAN 315
OY 309 GNIAFYNGFNRVAVDTGSNNMSEVLPOQETARTIIFNGKDLNVERIAVNPSPDPL-- 366
DB 316 STVASALDTNGTSATTELTLSKKIYP-----SPKVAIDQDKITLHDEKTLAPSKHQPTIS 371
OY 367 -ETTRKPDMTLK--EALKIAGFNEP-----NGNLQYQKGDITEFDFNEFDQOTSNI 414
DB 372 EQKMKEDADLKRMEILKSPHLKSPADRPQGRNRNSRNPSTDEETTKLAFIVEGQEN- 430
OY 415 KNOQLAELNATNIYVLDIKIKLAKNIIILIRDKRPHYDRNNIAVGADESVEKRAHREVINS 474
DB 431 -----NYNSTSRST-----EKKNDNTSAKKN--NGENKKKGKRPPELMSSTEAHAKVTE 478
OY 475 STEGLNLINDIRKILSGYIVEIEDEGELKEVINDRYDMLNI--SSLRQDGKTFIDPKK 532
DB 479 ETTKOIQSVRLDGRKVL-----QKVG-ESHIDSRNNTLNVPSKRPQGLGELPNPKK 530
OY 533 Y--NDKLPYISNPYKYNVAVYATKE 556
DB 531 HKPNEGRTPNISNGTINIQKLEPRE 556

```

DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE RETICULOCYTE BINDING PROTEIN 2 (FRAGMENT).
 GN RBP2.
 OS Plasmodium vivax (strain Belen).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN NCBI_TaxID=31273;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92315338; PubMed=1617731;
 RA Gallinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
 RT "A reticulocyte-binding protein complex of Plasmodium vivax
 mezozoites.";
 RL Cell 69:1213-1226(1992).
 CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
 CC HUMAN RETICULOCYTE CELLS.
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND (PROBABLE).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M88098; AAA29744.1; -
 KM Malaria: Receptor; Membrane.
 FT NON_TER 1
 FT NON_TER 1251
 SQ SEQUENCE 1251 AA; 143741 MW; 54BA51C7404AC572 CRC64;

Query Match 4.8%; Score 138.5; DB 1; Length 1251;
 Best Local Similarity 21.8%; Pred. No. 3.1;
 Matches 127; Conservative 96; Mismatches 227; Indels 133; Gaps 31;

QY 7 ELKQSSNRKKRSTASGTPDPRD-----NDGIPDSLEVEGT--VDVKNK----- 51
 DB 756 ELKFESEKNVNVLETENMSKNTNELDVHKNIDQYKVALEIDLSDELIDKQSSKLIEM 815
 QY 52 --RTPLSPWISNIHEKKGLTKYKSSPE---KMTASDPYDFEKYTGRIKVNVSPEARH 105
 DB 816 GMDIYIAKVLINQYKMK-SSIKSKEAVSVKIGNVSKKHSELSKIT-----CSQKSD 868
 QY 106 PLYAVAPYIVHDMENILISKNEQSTQNTDSETRTISKNTSRTHTSEVHGAENVANT 165
 DB 869 NITIALEK--QTELOHNLNRSFTQEKTNNSDSKLEIKITDFESLKNALKTLEG--EYNALK 924
 QY 166 STSRHTSEVHNAEVAHAVIDHSLSLAGEPTMAETFMGINTADTALNINIRVNGTAP 225
 DB 925 ASSDNH-EHVQSKSE---PVNPALSEI-EKEETIDISLNTA---LDELKKGRCEVS 974
 QY 226 IYNNVL--PTTSVLGKNQTLATIKAKENQLSQILAPNNYPSKNIAPIALNAODDSSTP 283
 DB 975 RKLIIDYTKELSDDELINTI---EKNVKAVLA---YIKKN-----YEDVYQDV 1019
 QY 284 ITMNTNQLELEKTKOLRIDTQOVGNIAITYNENGRVAVDGSNWEVLPQIQETTTARI 343
 DB 1020 LRLN-----EHNPTKV-----SNHEPTNF-----DKSNKSEELTKAVTDSKIT 1059
 QY 344 INGGDLNIVERRIAIVNPSDPLETKPMTLKEALKIAFGFENRGNIOYQCKDTEED 403
 DB 1060 ISKLKV-----ILEVNTENTMNTLESSAKETLEA---TNE---LKNKTSLSNET- 1103
 QY 404 FNFDOOTSONIKNOIALNA-----TNIY-TVLDRIKILAKNANILIRDKRFHYDRNNTA 456
 DB 1104 ----YQTSNEVKLOEKMSNADKYIDVSKIFNYLDTQKSIYVN-----QHSINN- 1150
 QY 457 VQADESVVKAHEVYVNSSTEGLLNIDDKIKIISGYVEIEDTGLEKVINDRYD--- 513
 DB 1151 ----KDKLKGKLOELIDADDSFTLESI-KKFNEIYSHIKTINIGELBOLQOTNKSEHDNVA 1205

QY 514 -----MLN-ISSLRODGKTFIDFKKYNDKPLIYSPNPK 547
 DB 1206 KHKKEIVHLINRVESLKGDKVKNH-DDDQYKKKILNASLNDNIK 1247

RESULT 7
 ID MYSL_YEAST STANDARD; PRT: 1928 AA.
 AC P08964;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DR MYOSIN-1 ISOFORM (TYPE II MYOSIN).
 GN MYO1 OR YHR023W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycaceae;
 OX NCBI_TaxID=4932;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=91088308; PubMed=2263482;
 RT Sweeney F.P., Watts F.Z., Pocklington M.J., Orr E.;
 RT "The MYO1 gene from Saccharomyces cerevisiae: its complete nucleotide
 RT sequence.";
 RL Nucleic Acids Res. 18:7147-7147(1990).
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=94378003; PubMed=8091229;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
 RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
 RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
 RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
 RA Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
 RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
 RA Vaudin M.;
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 RT VIII.";
 RL Science 265:2077-2082(1994).
 RN
 RP SEQUENCE OF 1-760 FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=88111539; PubMed=3322809;
 RA Watts F.Z., Shields G., Orr E.;
 RT "The yeast MYO1 gene encoding a myosin-like protein required for cell
 RT division.";
 RL EMBO J. 6:3499-3505(1987).
 CC -!- FUNCTION: REQUIRED FOR CELL DIVISION.
 CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X53947; CAA37894.1; -
 DR EMBL: X06187; CAA29550.1; -
 DR EMBL: 010399; AAB68872.1; -
 DR PIR: S05806; S05806.
 DR PIR: S12323; S12323.
 DR PIR: S46773; S46773.
 DR HSSP: P08799; 1MND.
 DR SGD: S0001065; MYO1.
 DR InterPro: IPR000048; IQ.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00063; myosin_head; 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR ProDom: PD000355; myosin_head; 1.

DR	SMART: SM00015; IQ: 1.	
DR	SMART: SM00242; MYSC: 1.	
DR	PROSITE: PS50096; IQ: 1.	
KW	Myosin: Actin-binding; ATP-binding; Coiled coil; Alkylation.	
FT	DOMAIN 1 793	MOsin HEAD-LINE.
FT	DOMAIN 794 823	IQ.
FT	DOMAIN 856 1911	COILED COIL (POTENTIAL).
FT	NP_BIND 180 187	AMP (BY SIMILARITY).
FT	DOMAIN 460 529	ACTIN-BINDING (BY SIMILARITY).
FT	MOD_RES 703 703	ALKYLATION (BY SIMILARITY).
FT	CONFLICT 36 36	K -> I (IN REF. 1).
FT	CONFLICT 46 46	I -> T (IN REF. 1 AND 3).
FT	CONFLICT 59 59	V -> S (IN REF. 1 AND 3).
FT	CONFLICT 86 86	L -> F (IN REF. 1).
FT	CONFLICT 330 330	MISSING (IN REF. 1 AND 3).
FT	CONFLICT 343 343	N -> S (IN REF. 1 AND 3).
FT	CONFLICT 421 425	COAKPT -> TLTSLI (IN REF. 1).
FT	CONFLICT 515 515	D -> S (IN REF. 1).
FT	CONFLICT 529 535	SKGPPG -> ARGHDR (IN REF. 1 AND 3).
FT	CONFLICT 541 541	D -> V (IN REF. 1 AND 3).
FT	CONFLICT 550 551	TD -> LM (IN REF. 1).
FT	CONFLICT 573 573	R -> A (IN REF. 1).
FT	CONFLICT 582 582	H -> D (IN REF. 1).
FT	CONFLICT 588 599	ETVSEWLSKKN -> NTLMKRGYPT (IN REF. 1).
FT	CONFLICT 599 599	MISSING (IN REF. 3).
FT	CONFLICT 627 632	EESSSA -> GNLLVC (IN REF. 1 AND 3).
FT	CONFLICT 695 695	R -> S (IN REF. 1 AND 3).
FT	CONFLICT 736 742	ENSTTPT -> RKPNNHD (IN REF. 3).
FT	CONFLICT 756 756	E -> R (IN REF. 1 AND 3).
FT	CONFLICT 773 784	NKLPKAGVLA -> ILTVQKLEYS (IN REF. 1).
FT	CONFLICT 793 794	KL -> NV (IN REF. 1).
FT	CONFLICT 896 896	N -> T (IN REF. 1).
FT	CONFLICT 900 900	N -> NSQITKNTNITETPOSTYIGERPKRYICGN (IN REF. 1).
FT	CONFLICT 906 906	N -> I (IN REF. 1).
FT	CONFLICT 911 911	N -> K (IN REF. 1).
FT	CONFLICT 915 930	NSGLNRAVTSSETILO -> RIATIKLPAINIT (IN REF. 1).
FT	CONFLICT 934 939	DDLJSE -> MTLFL (IN REF. 1).
FT	CONFLICT 951 953	AON -> RRI (IN REF. 1).
FT	CONFLICT 955 958	EEAH -> KKLD (IN REF. 1).
FT	CONFLICT 1002 1002	S -> C (IN REF. 1).
FT	CONFLICT 1049 1049	L -> D (IN REF. 1).
FT	CONFLICT 1056 1056	C -> S (IN REF. 1).
FT	CONFLICT 1060 1060	M -> I (IN REF. 1).
FT	CONFLICT 1085 1085	V -> E (IN REF. 1).
FT	CONFLICT 1123 1123	A -> C (IN REF. 1).
FT	CONFLICT 1143 1143	L -> S (IN REF. 1).
FT	CONFLICT 1144 1146	KSNI -> NLI (IN REF. 1).
FT	CONFLICT 1159 1168	REKREDOKK -> TRKREDOKE (IN REF. 1).
FT	CONFLICT 1179 1181	SKT -> ELKV (IN REF. 1).
FT	CONFLICT 1184 1185	LE -> WK (IN REF. 1).
FT	CONFLICT 1188 1204	LSOETSLNOLYKRRISG -> CHRRYKSLIKOKNIR (IN REF. 1).
FT	CONFLICT 1224 1224	P -> S (IN REF. 1).
FT	CONFLICT 1228 1228	E -> O (IN REF. 1).
FT	CONFLICT 1253 1253	E -> O (IN REF. 1).
FT	CONFLICT 1311 1323	POKESDINKLME -> LTKSLITNGNAS (IN REF. 1).
FT	CONFLICT 1400 1400	D -> H (IN REF. 1).
FT	CONFLICT 1454 1554	SEDLRFLDKLDESTEREKLLSSITKOKROOFENCMDDLOG NEKRLREHIDHLOAEEDVKNASITIEKLRONKOKELIM EREMENSDHLOLOETILE -> P (IN REF. 1).
FT	CONFLICT 1568 1568	D -> V (IN REF. 1).
FT	CONFLICT 1630 1646	DLKQDHTKVEEMLN -> SEARSLYKSGAVD (IN REF. 1).
FT	CONFLICT 1698 1704	MISSING (IN REF. 1).
FT	CONFLICT 1737 1737	TILOLOEONSUNG -> NTTANGTKFEM (IN REF. 1).
FT	CONFLICT 1754 1757	FODE -> LMM (IN REF. 1).
FT	CONFLICT 1777 1777	D -> E (IN REF. 1).
FT	CONFLICT 1788 1788	R -> T (IN REF. 1).
FT	CONFLICT 1825 1825	S -> D (IN REF. 1).

FT	CONFLICT	1882	1882	S -> W (IN REF. 1)
FT	CONFID	1902	1904	FWK -> NSGKRDLADD (IN REF. 1)
SO	SEQUENCE	1928 AA;	223634 MM;	6F54C701143D3C9F CRC64;
	Query Match	4.8%;	Score 138.5;	DB 1; Length 1928;
	Best Local Similarity	18.8%;	Pred. No. 5.7;	
	Matches 118;	Conservative 114;	Mismatches 219;	Indels 177; Gaps 28;
Qy	4	OLPELKQSSNSRRKKRSAGPTVVDNRDNGIDPSDELEVBG-YTVVYKXKRTFLSPWISNI	62	
Db	1097	ELLKKEKIDHKKKLAFFS---KQD---DAVSEHGKTTALKEKTRI-----	1138	
Qy	63	HEKKGLTYKKSPEKSWSPADSPDFEYKVGTRGDKNVSPEARHPVAAVPIVHVMENII	122	
Db	1139	---QLTEKSNYQK---IKREYSNFORET---KEQDOKRNSLV-----ESLN	1177	
Qy	123	LKSNEDOSTQNTDSEFTPTISKNTSTSRHTSEVHGNAEVNANTSTSRHTS-----	173	
Db	1178	DSKIELEAF-----LSOEISLNQYLNKRISGNS-VETNISTSTRSTSYSDPDLKE	1228	
Qy	174	-----EVHGNAEVHNAVIDSHSLAEERWAEF-MGLNTPADTRKLANIRY	218	
Db	1229	DIKKYVYDQLAETETIRNLENE---IEKKKLIRLPETETRIASSSEFDOKIYAKOMK	1285	
Qy	219	VNTGTAPIYNNVLPPTSLVGLKQNTLATIKAKENQSLQILAPNNYPSKRLAPIALNAODD	278	
Db	1286	LKKLIQDDMDSPIDPSIL--NEPLDNCPRDKSDINKMLLEVYDL--KRQDLETTRAHD	1340	
Qy	279	FSSPTITANNYNQLELEKTKQLRLDIDVYGINATINPENGRAVRADDTGSNMSEVLPQIE	338	
Db	1341	-AENAIISALHSKFRKIQ---GESSLSDDIY---KLKFEASEERKSLDEDKTKTTPLDNR	1393	
Qy	339	T---TARITFGKQDNLNVERIAAANPSPDLETPTKPDMLTKALKIAGFNEPNNGLY-	394	
Db	1394	TNLPYGDILKNNDSISKKEEIRYKK---LENYK---LQEL-----NESNKLQOL	1439	
Qy	395	-----QCKDITFEPDNFDQOTSQNIKNOL-----	418	
Db	1440	TLDLRQSKSEKALSEQDRLOKDESTEROKELLSTTKQOKQOFENCMQDLOGNEURL	1499	
Qy	419	-----AELNATNITYVLQDKIKL-NAKNMIIILRDKRFHYNDRNNIAVGADESYYKKAH	468	
Db	1500	REHIALKQAEEDVYNNMASIIEKTKQKQKREKLIMERNEM--ERNDSMQQLOETILE---	1554	
Qy	469	REVYNSTEGGLLNDIKDKITKLSGYIVE-----IEDTEGLKEVINDRYDMLNISSLR	521	
Db	1555	-----LKRQVQYKRLISDGLAHLEKRLSAVEDRSQYDEINRLKEELN-CSLK	1601	
Qy	522	QDGKTFIDEKKYNKDLPLXISNPNKVN	549	
Db	1602	AETNLKKEFAFLKYKLETTSTNDESAKIS	1629	
RESULT 8				
FAT1_SCHPO STANDARD: PRT: 1385 AA.				
ID	FAT1_SCHPO	013735; Q9JUT0;		
AC	013735; Q9JUT0;	15-JUL-1998 (Rel. 36, Created)		
DT	20-AUG-2001	(Rel. 40, Last sequence update)		
DT	20-AUG-2001	(Rel. 40, Last annotation update)		
DE	ACTIN INTERACTING PROTEIN 3 HOMOLOG.			
CN	FAT1 OR SPAC15A10.15 OR SPAC15E1.01.			
OS	Schizosaccharomyces pombe (fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC	Schizosaccharomycetes.			
OX	NCBI_TaxID=4896;			
BN	[1]			
RP	SEQUENCE OF 1-1033 FROM N.A.			
KC	STRAIN=972;			
RA	Murphy L., Harris D., Wood V., Bartell B.G., Rajadream M.A.;			
RL	submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.			

FT CONFLICT 1772 1772 D -> DEEDDE (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4B18 CRC64;

Query Match 4.7%; Score 136; DB 1; Length 1790;
Best Local Similarity 18.2%; Pred. No. 7; Indels 150; Gaps 25;
Matches 104; Conservative 114; Mismatches 204;

QY 2 NLOJPELKOKSSNSRRKSTSGAPTVPDRNDGIPDSLE-VEGYTVDVK---NKRRTPLSP 57
DB 1229 NLOJPELKOKSSNSRRKSTSGAPTVPDRNDGIPDSLE-VEGYTVDVK---NKRRTPLSP 1272
QY 58 WISNHEKKGGLTKYKSSPEKWTASDPSPDFKVTGRIDKNVSPARHPVLAAYIVAVHD 117
DB 1273 EYSELED-----KLKASEDKNSKYLELOKESKIKEEDDAKTE-----LKIQ 1315
QY 118 MENIT-LSKNEDOSTQNTDSETRTSKNTSTRTHTSEVHNAEVAHANTSTRTHTSEVH 176
DB 1316 LEKINTNLKAKKES-----ESELRLKKTSSSEERK-----LEDELI 1395
QY 177 GNAEVAHAVIDHSLSLAGEERTWAEWTGINTADTARLMANIRVYNTGTAPIYVNLPTTSLV 236
DB 1346 -NAEQLEKLEKKEIQIKKQAFERKLLNEGSSITTOEYSEKINT-----LEDELI 1395
QY 237 LGKNOTLATIRAKE-----NOLSOILAPNNYPPSKNLAPIALNADDDSS--TPTMTN 289
DB 1396 RLQNN--ELAKKEIDNRSELEKYSLSNDELLEKONTK--SIODELISYKDKITRDE 1452
QY 290 QLELEKTKQLRLDQVYGNATYVNFENGVRVDTGNSMSEVLPOJOETPAR---ITFN 346
DB 1453 KILSTERONKRLDESLEK-----EOLRAOESKAKVEGGLKLEESSEKAELEK 1502
QY 347 GKDL-----NLVERIAAVNPSDPLET--TKPDMLKEALKIAGFNEPNGNLQYOGKDT 400
DB 1503 SKEMAKLESTESNETELKS--METIRKSDKLEQSKSA---EEDIKNLQHEKSDLI 1557
QY 401 EEDFDFDOTSGNINKNOL-----AELNAT--NIYVLDRKILKAKNMLIRKRRHFD 451
DB 1558 S-RINESKDIELKSKRIEAKSSSELETYKQELNNAQKRIAAEENTVLKSK----- 1611
QY 452 RNNIIVGADSEVYKAEHREVINSTEGLLINDIKRILSGYIEIED-----TEG 503
DB 1612 -----LEDIERELKQKAE-----IKSNOBEKELLTRLELDQELDSTQOKAK 1656
QY 504 LKEVINDRYDMLNISSLRQDGKTIFDEPKYND 535
DB 1657 SEERRAEVRKFOYEKSOLEKAMLETKYND 1688

RESULT 10
YJL3_YEAST STANDARD; PRT: 1803 AA.
AC P47024; PB7192;
DT 01-FEB-1996 (Rel. 33, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSPOSON T4 207.7 KDA HYPOTHETICAL PROTEIN.
GN TYAB OR YJL113W OR J0780.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / FY1679;
RX MEDLINE=97103775; PubMed=8948101;
RA Czaplewski C., Kordes E., Pujol A., Jauniaux J.-C.;
RT "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X
RT reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,
RT SPI10, GCD14, REP1, PHO86, NCA3, ASF1, CCT7, GZF3, two tRNA genes,
RT three remnant delta elements and a Ty4 transposon.";
RT Yeast 12:1471-1474(1996).
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: 249389; CAAB9409.1; -
DR SGD: S0003649; YJL113W.
DR InterPro: IPR001584; Rye.
DR InterPro: IPR001878; ZnF_CCHC.
DR Pfam: PF00665; rve; 1.
DR SMART: SM00343; ZnF_C2HC; 1.
KW Transposable element; Hypothetical protein.
SQ SEQUENCE 1803 AA; 207693 MW; 16DCD72844BD52D3 CRC64;

Query Match 4.7%; Score 136; DB 1; Length 1803;
Best Local Similarity 20.5%; Pred. No. 7.1;
Matches 138; Conservative 91; Mismatches 247; Indels 198; Gaps 33;

QY 1 DMLQPELKOKSSNSRRKSTSGAPTVPDRNDGIPDSLEVEGYTVDVKNKRRTPLSPMIS 60
DB 1168 DONNITKELKYVDKNVPTDNGTVSPR-----LEQNTESGSPVQYVNSAFPLKERS 1219
QY 61 NIHERK-----GLTKY-----KSPKWTASDPSPDFKVTGRIDK 97
DB 1220 SLNMRKRKRHRDKNNSLSYELERDKKSKKNRVALLIDNMETVSAPIRIAYNEALSK 1279
QY 98 NVSPARHPVLAAYIVAVHDMENTILSKNEDOSTQNTDSETRTSK--TSTSRHTSEV 155
DB 1280 NDELEKEHYKQAY--HKEQLNLKMKYFDVDVYRSSE---IPDNILVPTNTTFTTKR 1333
QY 156 HG--NAEVAHANTSTRTHTSEVHNAEVAHAVIDHSLSLAGEERT--WAETMGINTADT--AR 211
DB 1334 NGTYKARIYCRDQTSPPDYSTTESLNHNHITKIFLMIANNRMFKTLIDINHFVLYAK 1393
QY 212 LMANIRVYNTGTAPIYVNL--TTSVLGKNOTLATIRAKENO-----LSQI-LAFNN 261
DB 1394 LEE-----IYHPHRCVVKLNKALYGLKQSPKRMNDHLQYLGILKNS 1443
QY 262 YYP-----SKNLAPIALNADDESSPTPTMYNOLF-ELEKTKDLR-----LDTD 305
DB 1444 YTPGLYQTEDEKMLM--IAVYVDCVIAASNEORLDEFINKLKNFELKITGTLIDVDLTD 1502
QY 306 QYVGNATYVNFENGVR-----RVDTGNSMSEVLPOJOETTARIIFNGKDLNLEVR 356
DB 1503 -ILGMDLYVNNKRLGITIDILKSFIRMD--KYNELKIKIRKSSIPHSYTK----- 1551
QY 357 IAAVNP--SDPLETTKPD-----MTIKEALKIAGFNEPNGNLQYOGKDTTDFNFDOOT 410
DB 1552 --IDPKKDVLOMSEERQGVGLKIQQL-----GELNYV--RHKCRDIFAVAK 1597
QY 411 SONITNOLAELNATYIVLVLDKIKLNAMNMLIRKRRHYDN-----NIAVGADESIVK 465
DB 1598 VARLVNYHERFYIMYIKIOLY-----VRYKDIGIHYDRCKDKKVIATDASVGS 1650
QY 466 EAHRE-----VI-----NSSTEGLLINDIKRILISGY----- 494
DB 1651 EYDAQSRIGVILIMYGMNIFNYSNKSTNRCSSTAELHAITEGTADETLKVTLEKEGE 1710
QY 495 -----IVEIEDP-----EGLK-----EVINDRYDMLNISSLRQDGKTFI-- 528
DB 1711 GDNNDIVMTDSKPAIQGLNRSYQOPKKEFTWIKRIEIIKEKIKESIKLKITGCGNIAD 1770
QY 529 -----DREKY 533
DB 1771 LTRKPVASDPKRF 1784

RESULT 11
YBV2_YEAST

ID YBV2_YEAST STANDARD; PRT; 753 AA.
AC P38261;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOHETICAL 85.5 KDA PROTEIN IN VPS15-YMC2 INTERGENIC REGION.
GN YBR102C OR YBR0831.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxId=4932;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=95208357; PubMed=7900426;
RA Mannhaupt G., Stucke R., Ehme S., Vetter I., Feldmann H.;
RT "Analysis of a 70 kb region on the right arm of yeast chromosome II.";
RL Yeast 10:11363-11381(1994).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: X78993; CAA55605.1; -
CC DR EMBL: Z35971; CAA85057.1; -
CC DR PIR: S44682; S44682.
CC DR SGD: S0000306; EXO84.
CC KM Hypothetical protein.
SO SEQUENCE 753 AA; 85537 MW; 1BCDE834466C3536 CRC64;

Query Match 4.7%; Score 134; DB 1; Length 753;
Best Local Similarity 19.9%; Pred. No. 2.8;
Matches 129; Conservative 93; Mismatches 227; Indels 198; Gaps 30;

QY 63 HEKGLT---KYSSPE---KSTASDPSDFE-----KVTGRIDKN 98
DB 15 HVKSSASPKQKTPPPAPKQKTKNPSDLDPATSYTLPIINARSRVATSMQRR 74
QY 99 VSPEARH--PLVAAYPIVHDMENIIISKNDOSTONTDS-----E 137
DB 75 LSHNTNYAPPTLDYSMPLEDMRMATVPNDNVSSHNSSSTTNESSYSGKPSNLS 134
QY 138 TRTISKNTS-----TSRTHSEVGHNAEVHANTSTSRTHSEVGHNAE--VH----- 182
DB 135 TADLSLNDSSYKNVPARSAMRNVTNPGSNDPFGNSTSLRKMILANPHFNKADFYHDLGN 194
QY 183 --AAVADHSLSLAGEKRWAEKMGINTADTARLANANITYVTGTPATYVNVLTPTSILVYGN 240
DB 195 ASAITTDKFTSNLTD-----LSIOVEEVKLANINKSYNEIMT--VNNDLVAMLEKLR- 245
QY 241 QTLATIKAKENOSQILAPNNYPSKNLA--PIALNAODEEST-----PITM 286
DB 246 -----VAANINDLNEVDQCTKAEKRLQLODOIDEROEGFNVEVSHNSPALLPKKA 300
QY 287 NYN-----OFLELEKTKOILRDLTDQVYGN-----ATYVFNFGVRVDTGSMSEVL 333
DB 301 GQGNGLMRDRSSVLLIEKFMDEL--DQLFKNVEGAQKFNSTKGHILMNSNMME-- 356
QY 334 PQIOETAR-----IIFNGDLNVERR-----IAAVNPSDPLETTPKDKTLEAL 379
DB 357 --LNTTGGKPLQWVOIFLNDLVLIADKSRKQNDPVSOCYPLKDVTVVQEEFSTK-- 411
QY 380 KIAFGNEPNCNLQYQCKDTEFD-----FNPDQOQSOMKKNQLAELNA 423
DB 412 RLIFKFSNSNSL-YECRADDECSRLLDVTRKAKDCLDIFHVEEENSKRIRSEFRYLOS 470
QY 424 T-----NIYTVLDKIKLNAKNNIILIRDKRFHYDRNNI 455

DB 471 TQQTGRENRRNSPNKRRSGSITPGRNVYTGAMDYLLQ-NLTLMSHRSRPRDMSST 529
QY 456 AVG---ADESVK---FAHREVINSSTEGLLNDKDKRLKLSGYIEDETEGLKEVIN 509
DB 530 AQRKLFDEGVEEIDIELARLFESAVETLL-----DIESOL-----EDLSERIS 574
QY 510 DRYAM-LNITSLRODGKFTIDFKKYNDKLPYISPNKVVVAVTK 555
DB 575 DEELMLNLISLKIQR--FAISSKLSOSIISNETHVHKSGTE 617
RESULT 12
P3K1_DICDI
ID P3K1_DICDI STANDARD; PRT; 1570 AA.
AC P54673;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PHOSPHATIDYLINOSITOL 3-KINASE 1 (EC 2.7.1.137) (PI3-KINASE)
DE (PBDINS-3-KINASE) (PI3K).
GN PIKA OR PIK1.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostellium.
OX NCBI_TaxId=44689;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=96009592; PubMed=7565716;
RA Zhou K., Takegawa K., Emr S.D., Firtel R.A.;
RT "A phosphatidylinositol (PI) kinase gene family in Dictyostelium
RT discoideum: biological roles of putative mammalian p110 and yeast
RT Vps34p PI 3-kinase homologs during growth and development.";
RL Mol. Cell. Biol. 15:5645-5656(1995).
CC -1- CATALYTIC ACTIVITY: ATP + 1-PHOSPHATIDYL-ID-MYO-INOSITOL = ADP +
CC 1-PHOSPHATIDYL-ID-MYO-INOSITOL 3-PHOSPHATE.
CC -1- SIMILARITY: BELONGS TO THE PI3/P14-KINASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: U23476; AAA85721.1; -
CC DR DictyDb: DD01099; PIKA.
DR InterPro: IPR002420; PI3K_C2.
DR InterPro: IPR000341; PI3K_ras_bld.
DR InterPro: IPR001263; PI3K.
DR InterPro: IPR000403; PI3_P14_kinase.
DR Pfam: PF00613; PI3Ka; 1.
DR Pfam: PF00792; PI3K_C2; 1.
DR Pfam: PF00794; PI3K_rbd; 1.
DR Pfam: PF00454; PI3_P14_kinase; 1.
DR SMART: SM00145; PI3Ka; 1.
DR SMART: SM00146; PI3K; 1.
DR SMART: SM00142; PI3K_rbd; 1.
DR SMART: SM00144; PI3K_P14; 1.
DR PROSITE: PS00915; PI3_4_KINASE_1; 1.
DR PROSITE: PS00916; PI3_4_KINASE_2; 1.
DR PROSITE: PS50290; PI3_4_KINASE_3; 1.
KW Transferrase; Kinase; Multigene family.
FT DOMAIN 41
FT DOMAIN 59
FT DOMAIN 73
FT DOMAIN 161
FT DOMAIN 294
FT DOMAIN 304
FT DOMAIN 308
FT DOMAIN 324
FT DOMAIN 413
FT DOMAIN 424
FT DOMAIN 503
FT DOMAIN 570
FT DOMAIN 821
FT DOMAIN 831
FT DOMAIN 836
POLY-ASN.
POLY-ASN.
POLY-ASN.
POLY-ASN.
POLY-THR.
POLY-THR.
POLY-ASN.
POLY-SER.
POLY-ASN.
POLY-ASN.
POLY-ASN.

FT DOMAIN 1309 1570 PI3K/PI4K
SQ SEQUENCE 1570 AA; 178374 MW; 558678872B34D783 CRC64;

Query Match 4.7%; Score 134; DB 1; Length 1570;
Best Local Similarity 21.4%; Pred. No. 7.5;
Matches 109; Conservative 68; Mismatches 205; Indels 128; Gaps 19;

OY 12 SSNSRKRKSTAGPTVPDRDNGIPDSLEVEGYVDVKNK-----RFLSPWISNIHEK 65
DB 176 SSSGSDSSASNOPIIRNRREGSILNLKQGLVKEISQRFQTPPTASTYTRNNANISIK 235
OY 66 KGLTYKSSPEKMWSTADPYDFEYVGTGRIDKNVSPERHPLVAAVPIVHVMEINILSK 125
DB 236 DKISLKEQERRKODSE-VQGRKXVIVLADSSNIQIYHPSVL-----IEKMSKLDLT 288
OY 126 NEDOSTQVTDSEPTISKNTSRTHTESEVHGNAEVHANTSTST-----RTHTESEVHGNAE 180
DB 289 EKKPATTTTTTTTTTSTSTSTSTPTTTTTT-----TNTSTNDITTKPKTSPKKNNEE 341
OY 181 -----VHAVALDHSLSLAGEKRTMAETMGL----- 204
DB 342 RSQSPITTPKOPVEEIVKVSSTPKS-NNTSKTSSDTPTGKTCKDKKKDKKSRDSCN 400
OY 205 -----NTADTARLNANIRVNTGTAPY-----NVLPTSLVIGKN-OPLATIKA-----K 249
DB 401 LVIYVNTNTSSNNNNNNNNNNNNNNETIIRKGRVLYVTPSSDLKKNIQIYFTPIIPVYK 460
OY 250 ENQISQILAPNNYPSKML---API-----ALNAODDESSPTITMANYNOFLEKTKQL 300
DB 461 TNKPNQLISNTSQOFLKTLISNEIPIDCKINDINDTDAFSDLSASASSSF--IRKSSQS 518
OY 301 RLDTQVYGNATYVPENGRVR-VDTGSWSEVLPQIOETTARIIFNGKDLNVERRIA 359
DB 519 LL-----NVQISLRKAIKTSFNILFLMPN-----QSKTKIQ 549
OY 360 VNPSPLETTKPDMLKEALIAFGE---NEPNCNLOYGCKDIPEFDPNQDSQNTK 415
DB 550 VKGSPTIE-----NLKERIISDYLFNNNSNNNNNNCKTGADSYLLIDFNDPMERSLV 603
OY 416 NQLAELNATNTYVLDKIKILAKNMILIRD 445
DB 604 NKSDYILDKRAQGLIPKLKVEIKSTILLSD 633

RESULT 13
SIRL_YEAST STANDARD: PRT; 678 AA.
AC P21691;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE REGULATOR PROTEIN SIR1 (SILENT INFORMATION REGULATOR 1).
GN SIR1 OR YKR101W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9117215; PubMed=2005909;
RA Stone E.M., Swanson M.J., Romeo A.M., Hicks J.B., Sternglanz R.:
RT "The SIR1 gene of Saccharomyces cerevisiae and its role as an
RT extragenic suppressor of several mating-defective mutants.";
RL Mol. Cell. Biol. 11:2253-2262(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Gallion L., DuJon B.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN REPRESSING TRANSCRIPTION OF THE SILENT
CC MATING-TYPE GENES, NAT1 AND ARD1, POSSIBLY VIA POSTTRANSLATIONAL
CC MODIFICATION OF OTHER SILENCING PROTEINS.
CC -1- PPM: IT IS POSSIBLE THAT N-TERMINAL ACETYLATION OF SIR1 IS

CC REQUIRED FOR COMPLETE SILENCING.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; 238524; AAA35046.1; -;
DR PIR; S14173; S14173.
DR SGI; S0001809; SIR1.
KW Transcription regulation; Repressor; Acetylation.
SQ SEQUENCE 678 AA; 79929 MW; 6540628CEFC2F3BC CRC64;

Query Match 4.6%; Score 132.5; DB 1; Length 678;
Best Local Similarity 20.8%; Pred. No. 2.9;
Matches 103; Conservative 78; Mismatches 195; Indels 119; Gaps 24;

OY 77 KWSTA--SDPYSDPEKVTGRI---DKNVSPEARHPLVAAVPIVH--VDMENIILSKNEQ 129
DB 142 KMGTYIHPMLKDKRKPLAGVCEFSUDVNPDRREYLI-EINSHQYIIIEGFLYLNERR 200
OY 130 STQNTDSETRTSKNTSTSRHTSEVHGNAEVHANTSTSTSRHTSEVHGNAEVHVAIDHS 189
DB 201 LYRWMDNNLR--SQGLTKMAHLRKTYNPVSLDIYSL-----NSNRYFVKDDL 248
OY 190 LSLAGERMAETMGINTADTARLNANIRVNTGTAPY--YNVLPTSLVIGKNQTLATJKA 248
DB 249 FOLLCKRVFVFKCVK-----MENGCKGKAPLIRVAKRTT-----TAKA 286
OY 249 KE--NQLSOLAPNNYPSKMLAPALNAODDESSPTITMANY--QFLELEKTK-OL 300
DB 287 THIAAIAINSTRPDSFKSKNNRYIVREKPIVENTISLMDSDIKKQOFTAEVYKKKI 346
OY 301 RLDTQVYGNATYVPE--ENGRVAVDTGSNMSEVLPQIOE---TARIIFNGKDLNV 353
DB 347 SADISQIENVHTQFNSQEKKNIRVKNVS--SEVLDQISKFPVSHVTLTLLMSAGODKNVI 404
OY 354 E-----RRIAAVNPSPLETTKPDMLKEALIAFGEFNEPNCNLOYGCKDITTEDFNPD 407
DB 405 ELVEELARLEKI---CIETKQSL---EELIDTFQAN-----PEMQASFD 444
OY 408 QOTSONIKNQLAELNATNTYVLDKIKILAKNMILIRPKRFHYDRNNTA---VGADESIV 464
DB 445 KEYYSIEEYKTTLELLIKEDLLITLIK--QMNMAAKKFTSEBYVSPRLVADGFLI 502
OY 465 KEAHREVINSSTEGILLNIDKIRKILSGYIEIDTEGLEKVINDRYDMLNSSLKODG 524
DB 503 DLAEKRPINPK-DPRLTLTLKDHORAM-----IDOMLVK----- 536
OY 525 KTFIDFKKYNDKPL 539
DB 537 --WPDFKRYDPIPL 549

RESULT 14
PCP1_SCHPO STANDARD: PRT; 1208 AA.
AC Q92351;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SPINDLE POLE BODY PROTEIN PCP1.
GN PCP1 OR SPAC69.06C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;


```

RN [1]
RP SEQUENCE FROM N.A.
RA Flory M.R., Morpheu M., Joseph J.D., Means A.R., Davis T.N.;
RT "Pcp1p, a Spo11p-related calmodulin target at the centrosome of the
RL fission yeast Schizosaccharomyces pombe."
RN [2]
RP Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN SEQUENCE FROM N.A.
RP STRAIN-972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Connor R.E.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: SPINDLE POLE BODY COMPONENT THAT BINDS CALMODULIN.
CC OVEREXPRESSION OF PCP1P CAUSES THE FORMATION OF SUPERNUMERARY SPB-
CC LIKE STRUCTURES AND DISRUPTS BOTH MITOTIC SPINDLE ASSEMBLY AND
CC CHROMOSOME SEGREGATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z81317; CAB03608.1; -
DR EMBL: AF348506; AAK31344.1; -
KM Coiled coil.
FT DOMAIN 151 375 COILED COIL (POTENTIAL).
FT DOMAIN 387 803 COILED COIL (POTENTIAL).
FT DOMAIN 874 1091 COILED COIL (POTENTIAL).
FT DOMAIN 1177 1204 COILED COIL (POTENTIAL).
SQ SEQUENCE 1208 AA; 140763 MW; 70264159ADDA42424 CRC64;

Query Match 4.68; Score 132; DB 1; Length 1208;
Best Local Similarity 20.0%; Pred. No. 6.7;
Matches 109; Conservative 95; Mismatches 232; Indels 108; Gaps 23;

OY 2 NLOI-----PELKQSSNRKRRKSTAGPYVPHDNDGIPRLSEGVTVYVKKRFELS 56
DB 203 NLOIQLKRTESLLOKSEKKNFKLEKY--DYLSKVND-VEGOSQVKKFTERR---FLE 255
OY 57 PWISNIHEKGLTFKYSKSPKSTASDPYSPFEKTCGRIDKNSPEARHPVAAVPIYHV 116
DB 256 NALEKVGREKDSLSTEMBEDSKNEKVEYERIQLONRUDE-LSEB-----L 301
OY 117 DMENIILSKNEDOSTQNTDSETRISKNTSTSRHTSEVHGNAEVHANTSTSRHTSEVH 176
DB 302 DVADQLLTERED-----EIAFL-KROIIEKENSASFENE--NSSYHLDE-- 345
OY 177 GNAEVHNAVDHSLSLAGERTMAETMGLNPTADTARLANA-IRYVNTGTAIYVNLPTTSL 235
DB 346 -----DYALILQKCEDEFADRIQVLTADLEKEKEMQIMHSEASIGLDSMOVHTL 395
OY 236 VLKGNQTLATIKAKENQSLQILAPNNYPSKNLAPIALNAODRSPPTWNYQFLE 295
DB 396 QEOUKANEELFELHDOISR-----NEEGNFEDIML-----QFSLDE 434
OY 296 KTYQRLDQVGNIAITYNFEENGVRVDTGSNMSEV-----LPQIQUETARIITFN 346
DB 435 E-----ERDVLESKLQTLIEDDNNLSRLMTSSLCNQIESLRTQNRREIDEEKNHLRLASK 488
OY 347 GKDLNLEERRAANVPNDPLETTKPRDMLKALKIAFGFNPNNGNLOXGQDITPEPFNF 406
DB 489 NSDALKALETRNLOEVRKELETLR--MKNSNDLNEIHDLREENGELTLKIDSTIKEDRL 546
OY 407 DOQTSQNIKN---OLAEINATNTIYVLDKIK-----LNAKNNIL-IRD---KRFHYDNN 454
DB 547 INLEQGIKSYEVAVSELNG-IDEYRNKRLKDKRETYNEVNAFOYKDNDRRRHESINK 605
OY 455 IAVGADSVYKAEHREYVNSTEGLLNIDK---IRKISGYVEIETDEGLKEVINDR 511
DB 606 LQDREKELTSLNLEKKNLVISSLRETVAMLEKERESIKKYISGNKKDLDT--NLMEIILDK 664

```

```

OY 512 YDML 515
DB 665 ISVL 668

RESULT 15
IPAA_SHIFL STANDARD; PRT; 633 AA.
AC IPAA_SHIFL
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE 70 KDA ANTIGEN.
GN IPAA.
OS Shigella flexneri.
OG Plasmid 210 kb invasion pWR100.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M90T / SEROTYPE 5;
RX MEDLINE=90221912; PubMed=2183200;
RA Venkatesan M.M., Buysse J.M.;
RT "Nucleotide sequence of invasion plasmid antigen gene ipaa from
RT Shigella flexneri 5";
RL Nucleic Acids Res. 18:1648-1648(1990).
RN [2]
RP SEQUENCE OF 1-88 FROM N.A.
RC STRAIN=M90T / SEROTYPE 5;
RX MEDLINE=69057927; PubMed=3057506;
RA Venkatesan M.M., Buysse J.M., Kopecko D.J.;
RT "Characterization of invasion plasmid antigen genes (ipABCD) from
RT Shigella flexneri.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9317-9321(1988).
CC - FUNCTION: ASSOCIATED WITH THE ENTRY OF THE BACTERIA INTO COLONIC
CC EPITHELIAL CELLS.
CC -----
CC - MISCELLANEOUS: SYNTHESIS OF THIS IMMUNOGEN IS RERESSED AT 30
CC DEGREES CELSIUS AND RESTORED AT 37 DEGREES CELSIUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17628; CA35624.1; -
DR EMBL: J04117; AAA26525.1; -
DR PIR: E31265; E31265.
DR PIR: S12763; S12763.
KM Antigen; Plasmid; Virulence.
SQ SEQUENCE 633 AA; 70093 MW; 2F804F45355E4751 CRC64;

Query Match 4.58; Score 130.5; DB 1; Length 633;
Best Local Similarity 21.3%; Pred. No. 3.4;
Matches 106; Conservative 68; Mismatches 193; Indels 131; Gaps 21;

OY 46 VDVKNRKF--LSPWISNIHEKGLTKYSSPEKWSSTASDPYSPFEKTCGRIDKNSPEA 103
DB 210 VDAQKLAFLDSLNRQKRTIDSKGFGIKSLRDLTAVFP-ELIRKVLNIDLEDIKDS- 267
OY 104 RHPVLAAPYIVHVMENIILSKNEDOSTQ-----NTDSETRISKNTSTSRHTSEVHG 157
DB 268 -HPIODGLPTPEEDMPOGPTPGANEKTSQVYIHINNDNRITDNRVFDNRVYDNSYHE 326
OY 158 NAEVHANTSTSRTH-TSEVHGNAEVH-----AAVDHSLSLAGERTMAETMGLNTA 207
DB 327 NPENDAOPTQOTDILLSRNGNSLILNFORALVQKVTVSLPHSIS----- 370

```


QY 181 YVEIGKILSRDILSKINOPYOKFLDVLTNTIKNASDSGODLLFTTNOLKEHPTDSVEFLE 240
 DB 181 YVEIGKILSRDILSKINOPYOKFLDVLTNTIKNASDSGODLLFTTNOLKEHPTDSVEFLE 240
 QY 241 QNSNEVOEFVAKAFYAYIEPQHRDVLQYAPAFVNMKFEQELNLSLEELKQDRLMSR 300
 DB 241 QNSNEVOEFVAKAFYAYIEPQHRDVLQYAPAFVNMKFEQELNLSLEELKQDRLMSR 300
 QY 301 YEKNEKIKOHYQHSNDSLSSEGRGLKLLKQPIPEPKKDDIHSLSQEEKLLKRLQIDSS 360
 DB 301 YEKNEKIKOHYQHSNDSLSSEGRGLKLLKQPIPEPKKDDIHSLSQEEKLLKRLQIDSS 360
 QY 361 DFLSTEEKEFKKLLQIDIRDSLSSEKELNRIQVDSNPLSEKEFKLLKLDIQPYD 420
 DB 361 DFLSTEEKEFKKLLQIDIRDSLSSEKELNRIQVDSNPLSEKEFKLLKLDIQPYD 420
 QY 421 INORLOPTGGLIDSPSINLDVRKQYKRDIONIDALLHOSGSTLYLNKTYLENMNNINLT 480
 DB 421 INORLOPTGGLIDSPSINLDVRKQYKRDIONIDALLHOSGSTLYLNKTYLENMNNINLT 480
 QY 481 ATTAGADVDSTDNTRKINGIFNEFKKNEKYSISSNMYIVDINERPALDNERLKMRIOLSP 540
 DB 481 ATTAGADVDSTDNTRKINGIFNEFKKNEKYSISSNMYIVDINERPALDNERLKMRIOLSP 540
 QY 541 DTRAGTYLNGKLLIQRNIGLEIKDVOILIKOSEKEYIRIDAKVVPKSIDTKIOEAQNLIN 600
 DB 541 DTRAGTYLNGKLLIQRNIGLEIKDVOILIKOSEKEYIRIDAKVVPKSIDTKIOEAQNLIN 600
 QY 601 QENKALGALPPTYKLTIFPNVHNRAVSINVESAYILINEMKNNISOLDKKATNTLVDSNG 660
 DB 601 QENKALGALPPTYKLTIFPNVHNRAVSINVESAYILINEMKNNISOLDKKATNTLVDSNG 660
 QY 661 RFVETDITLPRNIAQYHODEIYEQVHSKGLYVESRSILHGPSKGYELANDSEGFHE 720
 DB 661 RFVETDITLPRNIAQYHODEIYEQVHSKGLYVESRSILHGPSKGYELANDSEGFHE 720
 QY 721 FGNAVVDYAGYLLDKNOSDLVTNSKAFIDIFKEEGSNLTSGRTNEAEFFAEARLMHST 780
 DB 721 FGNAVVDYAGYLLDKNOSDLVTNSKAFIDIFKEEGSNLTSGRTNEAEFFAEARLMHST 780
 QY 781 DHAERLVOKNAPKTFQINDOIKETIINS 809
 DB 781 DHAERLVOKNAPKTFQINDOIKETIINS 809
 RESULT 2
 B59106
 hypothetical protein pXOI-122 - Bacillus anthracis virulence plasmid pXOI
 C:Species: Bacillus anthracis
 C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 09-Jun-2000
 C:Accession: B59106
 R:Okinkwa, R.T.; Clout, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Kelm, P.; Koehler
 J. Bacteriol. 181, 6509-6515, 1999
 A:Title: Sequence and organization of pXOI, the large Bacillus anthracis plasmid harbori
 A:Reference number: A59091; M0ID:99445483
 A:Accession: B59106
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-800 <OKI>
 A:Cross-references: GB:AF065404; NID:94894216; PIDN:AAD3426.1; PID:94894338
 A:Experimental source: strain Sterne
 A:Note: similar to calmodulin sensitive adenylate cyclase, edema factor, cya, plasmid pX
 C:Genetics:
 A:Gene: pXOI-122
 A:Genome: plasmid
 C:Superfamily: calmodulin-sensitive adenylate cyclase; calmodulin-sensitive adenylate cy
 F:34-286/Domain: lethal factor amino-terminal homology <LFA>

Query Match 12.2% Score 504; DB 2; Length 800;
 Best local similarity 24.1% Pred. No. 3.3e-13;
 Matches 214; Conservative 154; Mismatches 297; Indels 224; Gaps 40;

QY 4 KKEFI-----KVISMCLVATITSGPVFIPLVQAGAGHGVDYGMHKEKKNKDKRDE 59
 DB 3 RNFETPNKFSIISVLLFAISSQAIEVNA-----MNEVYTSDIKRNHKT 49
 QY 60 ERNKTOEHLAEIMKHIVIEVGEAAVKAERKLEKVPDVLVEMAKGIKTYIDG 119
 DB 50 EKNKTEKFKFDSJNNLVKTEFTNETLTKIQOTDOLKKIKFDVLEIYSELGELYTFDI 109
 QY 120 DITKHSLEALSDEKKKTKDIDYKDALHEHYVAKEGYEPVLVQSEDYVENTEKLN 179
 DB 110 DLVEHKLQDLSSEKNSMNSRGEKVPASRFRVEKKRETKLTI-NIKDAVINSQESKE 168
 QY 180 VVEIGKILSRDILSKINOPYOKFLD--VLNTIKNAS--DSGODLLFTTNOLKE--HPT 232
 DB 169 VVEIGKILSRDILSKINOPYOKFLD--VLNTIKNAS--DSGODLLFTTNOLKE--HPT 232
 QY 223 DPSVEFLQNSNEVOEFVAKAFYAYIEPQHRDVLQYAPAFVNMKFEQELNLSLEEL 292
 DB 224 SIDINFIKENLTERFOHAFSLAFSYFADHRTVLELYAPDMFEYMNKL----- 271
 QY 293 KDQRLMSYER--WEKIKOHYQHSNDSLSSEGRGLKLLKQPIPEPKKDDIHSLSQEEK 350
 DB 272 -----EKGEFEKI-----SESLKEG-----VERKIDVL-----KGEKA 301
 QY 351 LKRLQIDSDPLSTEEKEFKKLLQIDIRDSLSSEKELNRIQVDSNPLSEKEFKL 410
 DB 302 L-----KAGGLYPEHADAKFT-----ARLNTYILFRPNKILATN-----LI 339
 QY 411 KKLIDIOPIYDINORLODTGGLIDSPSINLDVRKQY-----KRDIONIDL-HQ-SIG 461
 DB 340 KSGVATKGLANVHGKSDPGVAGYIPFDQDLSKHGQDLAVEKNLENKKSITHEGEGIG 399
 QY 462 STLNKTYL-YENNININLTITLQADLDVSTDNTRKINGIFNEFKKNEK-----YSISSN 515
 DB 400 -----KIPKLKDLHRIEEL-----ENGIIKKGKKEIDNGKRYLLESN 438
 QY 516 YMIVDINERPALDNERLKMRIOLSPDTRAGYLE-NGKLIQORNGLEIKDVO-LIKOSEK 573
 DB 439 NQVYEF-RISDENNEVOYK-----IKEGKIYLVGEFNMKNRIEVAKNVAGYKLPETA 490
 QY 574 EYIRIDAKVVPKSIDTKIOEAQNLINOEMKALGLPKYTK-----LTFVNVHNRYA 625
 DB 491 DY---DLFALAPSLTEIKKQIPQ---KEMDKVNTPSLEKQGYTNLLIKYGIERK-- 541
 QY 626 SNIVESAVLLINEMKNNISOLDKKATNTLVDSNGRVPFDITLPRNIAQYHODEIY 683
 DB 542 ---PDSITKGTLSNMOKOMDLRL-----NEAVKYTYGTYGGDVNHNCTEDONEFP 587
 QY 684 EOVHSGKLYVESRSILHGSKGLYELANDSEGFHEFGHGVDDYA---GYL----- 732
 DB 588 EKQDEIFILNDEGEFILTNNEMTGRTLEKNITGKDLIYFNRSYNKLIAEGNKAYIEWTD 647
 QY 733 -LDKNOSDLVTNSKKEFI-----DIFKEEGSN-----LTSYGR 764
 DB 648 PITKAKINTIPTSAEFITKNLSIRSSNVGKYKSGDKDEPAKESYKTIAGYLSDYNS 707
 QY 765 NEAEFFAE-----APRLMSTDAER-LKVQKNAK-TPQINDOI 803
 DB 708 ANHIFSOEKKRKISIFRGIQAYNEIENVLKSKQIAPYKYNFOYLKERI 756
 RESULT 3
 JS0029
 adenylate cyclase (EC 4.6.1.1) precursor, calmodulin-sensitive - Bacillus anthracis
 N:Alternate names: anthrax toxin edema factor
 C:Species: Bacillus anthracis
 C:Date: 31-Mar-1992 #sequence_revision 11-Nov-1994 #text_change 19-Jan-2001
 C:Accession: JS0029; PS0307; JS0602
 R:Robertson, D.L.; Tippetts, M.T.; Leppia, S.H.
 Gene 73, 363-371, 1988
 A:Title: Nucleotide sequence of the Bacillus anthracis edema factor
 A:Reference number: JS0029; M0ID:89211974
 A:Accession: JS0029

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:46:05 ; Search time 72.22 Seconds

(without alignments)
853.298 Million cell updates/sec

Title: US-09-747-521-2

Perfect score: 4145

Sequence: 1 MNIRKEFIKIVMSCLVTAT.....KNAPKTFQFIINDIKFIINS 809

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR68:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	4145	100.0	809	1 J00032	anthrax toxin lethal
2	504	12.2	800	1 B59106	hypothetical prote
3	497	12.0	800	1 J50029	adenylate cyclase
4	264.5	6.4	1979	2 C71622	hypothetical prote
5	263.5	6.4	2401	2 T28676	hypothetical prote
6	255	6.2	2269	2 T28677	hypothetical prote
7	246.5	5.9	1939	2 T18372	repeat organellar
8	245.5	5.9	1127	2 T28317	ORF MSV156 hypothe
9	241	5.8	2829	2 A42771	reticulocyte-bind
10	239.5	5.8	3119	2 T18414	protein g377 - mal
11	235.5	5.7	2663	1 S28261	centromere protein
12	234	5.6	946	2 S28061	SCP1 protein - rat
13	230	5.5	993	2 S49461	synaptonemal compl
14	229	5.5	3724	2 T18427	hypothetical prote
15	224.5	5.4	1558	2 B71603	RESA-H3 antigen pr
16	222.5	5.4	2166	2 G70163	hypothetical prote
17	218.5	5.3	1163	2 D63115	type I restriction
18	218	5.3	980	2 C64236	protein V (fcrv) h
19	218	5.3	756	2 E71606	hypothetical prote
20	217	5.2	1088	2 T18559	hypothetical prote
21	216.5	5.2	1679	2 S48385	hypothetical prote
22	216	5.2	1005	2 A64465	hypothetical prote
23	215	5.2	2245	2 T18278	myosin heavy chain
24	214.5	5.2	1313	2 F96673	hypothetical prote
25	214	5.2	3225	2 I52300	giantin - human
26	214	5.2	3259	1 A56539	giantin - human
27	213	5.1	1516	2 E71619	RAD2 endonuclease
28	212.5	5.1	1957	2 T38077	hypothetical colle
29	211.5	5.1	1875	2 S38173	myosin-like protel

30	211	5.1	1252	2 B42771	reticulocyte-bind
31	210	5.1	1939	2 I48175	myosin heavy chain
32	209.5	5.1	1128	2 G86266	hypothetical prote
33	209.5	5.1	1780	2 T17272	hypothetical prote
34	209	5.0	1642	2 T08880	NMDA receptor-bind
35	208.5	5.0	691	2 S48390	hypothetical prote
36	208.5	5.0	841	2 A86188	hypothetical prote
37	208	5.0	1230	2 S56850	SMC1 protein homol
38	208	5.0	2710	2 A37052	toxlin A - Clostrid
39	207	5.0	1922	2 T00637	hypothetical prote
40	207	5.0	3394	2 T18501	hypothetical prote
41	206.5	5.0	1650	2 T18444	hypothetical prote
42	206.5	5.0	1837	2 T41023	probable nuclear p
43	206	5.0	2139	2 T18296	myosin heavy chain
44	205.5	5.0	1738	2 T14867	interactin - slime
45	204	4.9	1312	1 BMYBDL	RAD50 protein - ye

ALIGNMENTS

RESULT 1

J00032 anthrax toxin lethal factor pXOI-107 precursor - Bacillus anthracis virulence plasmid

C:Species: Bacillus anthracis

C:Date: 31-Mar-1990 #sequence_revision 11-Nov-1994 #text_change 11-May-2000

C:Accession: J00032; C59104

B:Bragg, T.S.; Robertson, D.L.

Gene 81, 45-54, 1989

A:Title: Nucleotide sequence and analysis of the lethal factor gene (lef) from Bacill

A:Reference number: J00032; MUID:90034185

A:Accession: J00032

A:Molecule type: DNA

A:Residues: 1-809 <BNA>

A:Cross-references: GB:M29081; NID:9143143; PIDN:AAAT9216.1; PID:9143144

R:Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koeh

B: Bacteriol. 181, 6509-6515, 1999

A:Title: Sequence and organization of pXOI, the large Bacillus anthracis plasmid harb

A:Reference number: A59091; MUID:99445483

A:Accession: C59104

A:Molecule type: DNA

A:Residues: 1-809 <OKT>

A:Cross-references: GB:AF065404; NID:94894216; PIDN:AAD32411.1; PID:94894323

A:Experimental source: strain Sterne

A:Note: Similar to Anthrax toxin lethal factor precursor; lef, plasmid pXOI, B. anthr

A:Comment: This lethal factor of Bacillus anthracis is part of the tripartite protein

but they cause anthrax, an infectious and often fatal disease of cattle, sheep, and o

C:Genetics:

A:Gene: lef; pXOI-107

A:Genome: plasmid

C:Superfamily: anthrax toxin lethal factor; lethal factor amino-terminal homology

C:Keywords: toxin

F:1-33/Domain: signal sequence #status predicted <SUS>

F:34-809/Product: anthrax toxin lethal factor #status predicted <MAT>

F:44-295/Domain: lethal factor amino-terminal homology <LFA>

Query Match 100.0%; Score 4145; DB 1; Length 809;

Best Local Similarity 100.0%; Pred. No. 3.1e-158; Indels 0; Gaps 0;

Matches 809; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MNIRKEFIKIVMSCLVTATLTS	GPVPIPLVQAGGHDVGMVKKKKDKKKRDEE	60
DB	1	MNIRKEFIKIVMSCLVTATLTS	GPVPIPLVQAGGHDVGMVKKKKDKKKRDEE	60
QY	61	RNKTQEHLEIKMKHYVKEVKG	EEAVKKAELLEKVPDYLEMKATGKTYIVDGD	120
DB	61	RNKTQEHLEIKMKHYVKEVKG	EEAVKKAELLEKVPDYLEMKATGKTYIVDGD	120
QY	121	ITRHISLEALSEDKKTKTKOY	GDALLHENYVAKGYPVLYIOSSEDEVTEKALNV	180
DB	121	ITRHISLEALSEDKKTKTKOY	GDALLHENYVAKGYPVLYIOSSEDEVTEKALNV	180

0

```
A:Molecule type: DNA
A:Residues: 1-800 <ROB>
A:Cross-references: GB:M24074; NID:g142812; PIDD:AAA79215.1; PID:g142813
A:Accession: F50307
A:Molecule type: protein
A:Residues: 34-48 <RO2>
R:Escuyer, V., J. Duflot, E., Sezer, O., Danchin, A., Mock, M.
Gene 71, 293-298, 1988
A:Title: Structural homology between virulence-associated bacterial adenylate cyclases.
A:Reference number: J50033; M0ID:89138004
A:Accession: J50602
A:Molecule type: DNA
A:Residues: 1-349, 'V', 351-509, 'Q', 511, 'EW', 514-800 <ESC>
A:Cross-references: GB:M23179; NID:g142814; PIDD:AAA22374.1; PID:g142815
A:Comment: This enzyme is activated by calmodulin and increases the intracellular cAMP c
C:Genetics:
A:Gene: cya
C:Superfamily: calmodulin-sensitive adenylate cyclase; calmodulin-sensitive adenylate cy
C:Keywords: nucleotide binding, P-loop; phosphotus-oxygen lyase; toxin
E:1-33/Domain: signal sequence #status predicted <SIG>
E:34-800/Product: adenylate cyclase, calmodulin-sensitive #status predicted <MAT>
E:34-286/Domain: lethal factor amino-terminal homology <LFA>
E:303-607/Domain: calmodulin-binding adenylate cyclase catalytic domain homology <ADBE
E:313-333/Region: calmodulin-binding #status predicted
E:347-354/Region: nucleotide-binding motif A (P-loop)
```

[illegible]

RESULT 5
T28676
Thoptry protein - Plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text-change 01-Dec-2000
C:Accession: T28676; A43521
R:Sluha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A:Title: Comparison of two members of a multigene family coding for high-molecular mass
A:Reference number: Z20507; MUID:97077455
A:Accession: T28676
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>
A:Cross-references: EMBL:U36927; NID:G1041784; PID:G1041785; PIDN:ABB41263.1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a Plasmodium yoelii thoptry protein. Multiple co
A:Reference number: A45521; MUID:91101660
A:Accession: A45521
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 2260-2401 <KEE>
A:Cross-references: GB:M34281

```

Query Match 6.4%; Score 263.5; DB 2: length 2401;
Best Local Similarity 20.3%; Pred No. 0.0051;
Matches 193; Conservative 168; Mismatches 358; Indels 233; Gaps 42;

QY 2 NIKKEFIKVISNCULVATITLSGPVPIPLVQAGHGHDVGMHKKKKENKDKRKRDEER 61
    ||| ||| : : : ||| : : ||
DB 1389 NIYEERIKSYDLITHYLETVFSKEPTIEQIKKRIATQNELLTINIKVNAKAKSYLDIEA 1448
    ||| ||| : : : ||| : : ||

QY 62 NKTOE-EHLKELMKHI-----VKTEYKGE-----AVKKEAEKLEKVPDVL 104
    ||| ||| : : : ||| : : ||
DB 1449 NEEDRIIVTHFFKRLANDVNDKFTNEYSKVN-KGPDNISNSINNVKKSTDELTLLINIQTK 1507
    ||| ||| : : : ||| : : ||

QY 105 EKKYKAIGCKIYI-----VDG-DITKHIS--LEALSEDK 134
    ||| ||| : : : ||| : : ||
DB 1508 EKNYIVSKRYIYKYEAENIFINIPKLANSLTIQIKSSGIDLEKININAILPYLDSOK 1567
    ||| ||| : : : ||| : : ||

```

```

01 135 KIKIYKODALLHHYAYAKEGEYPLV-VLOSSEDVYENKELNAYNYELGILISDIL 193
02 1568 KDTLFIISPEKSTSTTYTKIKIDSINTYLLDILKROELOKKOQOLNLF-----NMLH 1622
03 194 SKINOPYKFLDVLNTIKINASDSOGODLLFTNOLEHPTDSVEFLBONSNOVEOYAKA 253
04 1623 DKV-OATNELKDITSLDKNKKEQ-----ILNKYV-----LLHKSNELNKLSCNS 1666
05 254 FAY--YIEPQHDVLOLYAPAFNYMOKFNQOEINLSLELKODRMLSTRYKMKIKOHY 311
06 1667 QNYDTILLESSKYDKIK--EKSNMYEKENGINFDVKAHEO-FNNIDIKIEKLENNY 1722
07 312 OHMSD--SLSEEGHGL--KILQ-----IPIEPK--KDIHLSOEEK-- 350
08 1723 KHSEKDNANFSEENNIIOSKKKLKELTNAFNAELKIEDKIEKNGILINKLETREDCM 1782
09 351 -----LLKRIQIDSSDF-----LSTEEKEFLKQIDIDISLSEEEKL-----LNRI 393
10 1783 LETYKTEVEPLKIKTYDTKFTSATKPSKREFLAYIDA-TSNSLNDIDINTLOQTYVDNQI 1841
11 394 Q-----VDSNPLSEKEKEPEFLKKLKDIDOPYDINOQLDTGCLIDSPSINLOVRK 443
12 1842 NKHVASWADATNDNNNLIEKEKEATKTINNLTELF-----TIDSNKIDADGLH 1890
13 444 QYKRIQIDALLHOSIGS--TYNKIYLYENMNI-----NNLTATLGLADL 487
14 1891 NKKIOLITFNSLHSDISIDIKOLYKMHAFKLLNGHINKKRYDISKEFDNILOLSESL 1950
15 488 VDSIDNTKINGIENEFKKNFKRYSISSNYMVLVDINERPALDERLKWRIOLSPTTRAGYL 547
16 1951 TANLNDLKEIGQIKSIDKKKQOFLHMLSEY-PIPNFTLKEIYHIDLYVQKKNQDELENTNE 2009
17 548 ENKGIILORNGLEL-----KDVQIKOSEKEYIRIDAKVYPRKSIDTKI 592
18 2010 ENENTITLVMDIITTKMKKVOISLNEVTTYENDSNIKIHOIDNNENNDVSKI-KESLETTI 2068
19 593 QEOALINOENKNKALGLRKYTKLTGFENVHNRKYASVNIYEOAVALLNEMKNINOSOLIKKYT 652
20 2069 OSFOKILN-----KLMEIKAQFOYDNNNNINNVISTISQDVNDVKKHHSKRL--TIE 2116
21 653 NYLVDNGRFVFTDITLPLNIAEOY--THODEIYEQVHSGKGLYVPESRSILHGFSGVEL 710
22 2117 NELLQ-----IQSLSEDIKKSTYDIRSEQITKYVNPIDHYVEQOYKKIIONBNK---- 2165
23 711 RNDSCGFHEGHAVDVTAGYLLDKN-QSOL-----VTRSK-----KFTDIRKEE 754
24 2166 -----DEIDDLIOEITVNNKSESLKPLTIINKKNKNVPIISRDKVINLIKSE 2213
25 755 GSN--LTSYGRTEAEFFAEAFRLMHSTDHAEPL--KVQCNAPKTPOFIND 801
26 2214 YNNNDNVSYNAKKE--EDANNITRIDTSHNMINDLQKN-----FKIIDO 2259

```

```

RESULT 6
128677
Rhoptry protein - Plasmodium yoelii
Species: Plasmodium yoelii
Date: 15-Oct-1999 #sequence-revision 15-Oct-1999 #text-change 01-Dec-2000
Accession: 128677; C45521
R. Keen, J.; Sinha, K.; Brown, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994
Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii
Reference number: Z20508; MUID:95021522
Accession: 128677
Status: preliminary; translated from GR/EMBL/DBJ
Molecule type: DNA
Residues: 1-2269 <KEE>
Cross-references: EMBL:127838; NID:9457145; PID:9457146; PDI:AAA21304.1
R. Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple
Reference number: A45521; MUID:91101660

```

A:Accession: C45521
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 2131-2269 <KE2>
A:Cross-references: GB:M34283

Query Match	6.28	Score 255	DB 2	Length 2269
Best Local Similarly	20.58	Pred. No. 0.01		
Matches 210	Conservative 155	Mismatches 307	Indels 350	Gaps 51

Qy	39	DVGHVHEKKNDEKRNKDE	-----ERNTOEHLKEIMK	-----HIYI	-----79		
Db	1187	EVSNIKINENYVQKKKNEI	IGIIVKEKINEIATNKNQ	LESTKELKPTIOTHISSFAN	12468		
Qy	80	EVKG	-----BEAVK-EAEKLEKPPSVLEMYAIGKIIYVGD		120		
Db	1247	DLEGIDSDENLGKYNTEBMGNI	EEPIKSYNLITNYLEVYSKESI	-TYNQIOKKRIDTQKE	1305		
Qy	121	ITKRIIS	-----LEALSDD	-----KKIKIYIKDALLHEHYIAKEGEPEV	161		
Db	1306	LLKNIENVNKAASYIDIK	ENEDRIYTHKKRLNTV	-NDNFKE-YSKVNEGDMJISN	13628		
Qy	162	-----	LVIOSSDY	-----VENT	-----EKALVYVEIGI	187	
Db	1363	SINTVKNSTDBNSILNTLNOT	KEMYANINYNTYISKYAEINFINRIKLAITMIKIKN		142228		
Qy	188	-----LSRDI	-----	-----	-----LSKINOPYOKFLDYLTINKNAS	214	
Db	1423	SSGIDLSKOIKIALI	SYLDSKTEDTLFI	FPSPQKTEYTKISDSYSILDIKSKSQELQ		14828	
Qy	215	DSDDQ	-----DLF	-----TNQKEHPD	-----SVEFLDONSNEVOEYRAK	252	
Db	1483	KKEOQTKLLFENRRLYEK	VOAENELRGTLSDIKYKREKILSEVKILLHKSNEMLKSCN		15428		
Qy	253	AFAYIEPOHRDYLOL	YAPAFNYMOKFNOEINLSLEELK	-----DOJMLSRKME	-----K	306	
Db	1543	F	-----QVNDTIL	-----ESSKY	-----DOYKEKSNNYQOEKIKGIDPENVTDMEEKNNNDIK	15908	
Qy	307	IKQHYQMSDLSSEEGGLLKQ	-----	-----PIEBK	-----KDDIHSLSQEEK	349	
Db	1591	VIEELENNYDS	-SEENNINIKOSKLEKELTNKFNAEIKKIDIKITEKNDLIDKLETERKN		164498		
Qy	350	-----	ELKRIODISSDFLSTE	-----EKPEFLKQIDIDIRDSISEBEKL	-----LN	391	
Db	1650	CMFLTHTTAAETLKIKIT	TDISKFTESKTRKSEFLYIG	-DYSNSINDIATVLOUKYDHL		17088	
Qy	392	RIQ	-----	-----VDSNPISEKEKEFLKLKIDIOPYDINORLODGTGLIDSPSINDV		441	
Db	1709	QINKYVTSKLSADATNONN	NILKEKE	-----	-----ATQAKNLKTLFIDSNMIDA	1758	
Qy	442	RKQYKRIQONI	-DALLHOSIGS	-----PLYNKIYIYEMNI	-----	476	
Db	1756	NALNNRNIOWYFENSELHKS	IESIKOLYKMHVKLLINTGOINGKYFYDISKOPNDILOQ		18158		
Qy	477	-NNLTATIGADL	-VDSDTMTKINRGI	-----	-----FNEPKKNFYSISSNYMYVDINE	523	
Db	1816	ESELTANLN-DLKEIGIKID	KKNKFLHALNETPPIFNFTLKEIHYIDIKYKRYQIDELEN		18748		
Qy	524	RPALDNE	-----	-----RLKWRIOUSPDRRAGYLENGKLILORNI	-----IGLEIKDYOIIKQSEK	573	
Db	1875	ITSEENENITLYIDTIT	KLKEKYOSILNFTYENDSNIITKHOIDOTNENDYSKIKESIK		193348		
Qy	574	EYINLDAKVPK	-----	-----SKIDTQIGEQOLINDEM	-----NKALGL	609	
Db	1935	TTIOSFOELIKKINGIK	AOFYDNNNNININNTISTISODVNDVKHHSISDILLTEHELLEI		19948		
Qy	610	PKYKFLI	-----TFNV	-----HNRYASINIVESA	-----VALLINENKKNNO	-----SDLKRTYNTNLVGN	659
Db	1995	QKSLIEDIKNSYELIRGANN	NNYVMTIRINYVBOQNKIQNNSNDEIDDIQIKLANRKE	-----	2058		
Qy	660	GRVFTDITLPI	NAEQYTHODEIYEOVHSKGLYVPSRSILHGPSKV	-----ELRND	713		

```

Db 2053 -----SEKLEPTINGKNNNNTTIIISRI- NKVINLIESE-----YGNNNNNVTYVAKKLEED 2102
QY 714 SEGFIERGHAUNDYAGYLDKROS---DLYTNSKFFDIK-----EESNLTXYGT 764
Db 2103 ANSTILD-----LDKSONILKDLIOONLKIIDLKKNKOEJENRNNOITINRE 2150
QY 765 NE 766
Db 2151 OE 2152

```

```

RESULT 7
T18372
repeat organellar protein - Plasmodium chabaudi
C:Species: Plasmodium chabaudi
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18372
R:Werner, E.B.; Taylor, W.R.; Holder, A.A.
Mol. Biochem. Parasitol. 94, 185-196, 1998
A:Title: A Plasmodium chabaudi protein contains a repetitive region with a predicted
A:Reference number: Z18922; MUID:98418765
A:Accession: T18372
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1939 <WEB>
A:Cross-References: EMBL:U43145; NID:g1151157; PID:g1151158; PIDN:AAC63403.1

```

Query Match	5.9%;	Score 246.5;	DB 2;	Length 1939;
Best Local Similarity	19.68;	Pred. No. 0.019;		
Matches 198;	Conservative 174;	Mismatches 302;	Indels 335;	Gaps 50;

QY	36	GHGV--GMHYKKEKKDKENKKDDEBNKTOZEHKLEIM-----KH---LYKIEVG--	83
Db	898	GHREVYAGLEKKHEVVALEEKHEKHIEALALEGCH--KEYVALEGKHREVAAGLEKAPNL	956
OY	84	EEAVKKEAAEKLEIKVPSDVL-----EWKALGKXIYIVGDITK-----HISLEALEDK	135
Db	957	EEGKHKEWAAE--LEKRRADIVALEOHKA-----ELIKLGEHKEVVAALIEEKY	1004
OY	136	KIKDIYKDALHHEHYAAKEGYEPVLVIQSSDEVVENTEKALNVEYIGKILSDILSK	195
Db	1005	KVEAL-----KLAEEH-----VVTKLGSQHKEETAKLEDGCHKEVAVNEVEKKASL	1051
OY	196	INQAPQFLVLYNTIKNASDSDODLLFTNQLKEHHTDSVFELEJONSNEVOGAPAKAFA	255
Db	1052	LNMLEEHNKNEMIKLEKHESADILEKLYQODEVKNNSNKTEILTVIKDL-----	1105
OY	256	YUIEPQHRDVLQALAPAFVNMDFNE--OEIN--LSLEELKD-----	294
Db	1106	-----NDSIMCKKQILDEVEKRNRYNEIKKLIVQNEKKDMNDKILEKENEIKL	1158
OY	295	QRMUSREKWEKIKOHYOH-----WSDSLSE--EGGGLLKQIPIE	334
Db	1159	NKRLSNKYVEFTKENTKYKSEMYVNEKKERIIYDSYCKENISDSDEGKGNLKMPTLSLK	1218
OY	335	PKKDIIHSLSQEEKELKLRIQDSSDPLSTEEKELKLOI-----	376
Db	1219	KKENITFSINDKNF-----SSELVYTISAYINKIEMKKEIEDNGKINIEDLNK	1265
OY	377	-----DIRDSLSEEEKELNLRIQVDSNPLSEK-----EPLK-----KULK	414
Db	1270	ILDSLNETLINENKKNVLTDEBNNLKKEIIE--KDNKLEKNEKNENETLNLMDITKLK	1328
OY	415	DIQPPDINQRLQDTQGLIDSPSINL-----DYKQYKRIQNIIDALHOSIG--STLYNK	467
Db	1329	EISEW-----KDEEEKLTKENIKLKNDIEQINKKEKRIEENMLIKFENINEVJSLNQ	1382
OY	468	IYL-----YE-----NMNI--NLTLATGADIVDSTDN--TKINRIGIFNE--	503
Db	1383	IEIEKMKLEELNKYTELLLAKEKRETNMSINDONKIVENNILEDDSKONMLNKAVEEKT	1442
OY	504	-----FKKNFYSSISSNYMIVDINERPALDNERLKWRIQLSPDTRAGYLENGKLLIQRNI	558

Db 1443 GDDINCEKNNDQAKKEISYLDKDELKIKTSMUGDEL-----NRKNKYDEKVK-----NL 1489
 QY 559 GLEIKVQV-----IKOSEK-----EYRID 579
 Db 1490 TMELELKIRNKKGEBAIAELNKLNIKKEKNSVKQDESSNNIITKDGDKTPEEVSND 1549
 QY 580 AKVVP--KSKIDPKIOBAQUNINOENNKALGDKPKYTKLTFNHNRRASNVESAVILN 637
 Db 1550 DKIQKRWKANLYLKEKP---DLMDNINSLEK-----ENFRVSTIVKE-----1550
 QY 638 EWMKNINOSDLIKKVTYLYVDNGNR-----FVFETDI--TLPIAIQYHODEIYEQVH 687
 Db 1591 --NRKNVQNDKIYGISYFPKCEKELKNDMLVICLYLKDILSILELNNFVNLFEKIDIL 1648
 QY 688 SKGLIYVP-ESRSLILHGPSKGVELRN---DSRGFT---HEGCHAV-----D 726
 Db 1649 WKQMYIPTETIRILEFRYFSLDKLRNVKCVNDEYVNNERYESWALFQTYLETASMLIK 1708
 QY 727 DVAGYLLDNOSDLVYNSK-----KFIDI-----FKREGSNLT 759
 Db 1709 EMITYYLEAEKEDSCENSNENFDPKPTITDLNFSKDISIRIKTTAQLKKEINFEARKNIL 1768
 QY 760 SYGRTEAEFEFAEAFRLIMHSTDAERLKYQKNAPKTFQF--INDQIKRTI 807
 Db 1769 NVDYQIILMKYHECLR-----KLKIYKNAARELDFYVNVSSKRSI 1808

 RESULT 8
 728317
 ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
 C:Species: Melanoplus sanguinipes entomopoxvirus
 C:Date: 21-Jan-2000 #sequence=revision 21-Jan-2000 #text_change 21-Jul-2000
 C:Accession: T28317
 R:Alonso, C.L.; Tufman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
 J:Virol. 73, 533-552, 1999
 A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
 A:Reference number: Z20484; MUID:99102612
 A:Accession: T28317
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1127 <AFO>
 A:Cross-references: EMBL:AF063866; NID:g40496647; PIDN:AAC97677.1; PID:g4049717
 C:Genetics:
 A:Note: MSV156

 Query Match 5.9%; Score 245.5; DB 2; Length 1127;
 Best Local Similarity 20.6%; Pred. No. 0.01;
 Matches 167; Conservative 126; Mismatches 263; Indels 253; Gaps 37

 QY 60 EENKTQEE--HLKETMKHIVKIEVGEBAVKEAEKLEKYPVDLEMYKKAIGKTY- 115
 Db 24 ENNKVSLDIINSYLETNNI-----KFSDKITNELIKYNNKIVKEKIFY 66
 QY 116 -----IYVGDITKHISLALSFDKKIKD-----IYGDALLHENYVYAKEGTEPV 161
 Db 67 MHNOEKINDYNIITLOYLIEYNNNEINCIREKNKPPCKNPV-----NITYKKKLIYD 118
 QY 162 IYIOSSEDEVNTEKRLANYVE---IGKLSRDILSKINQPY-----QKFLDYV 207
 Db 119 L-----DYEEKKDKELVINIEQKNAVDKI--NDIKNNVNIHSDNETIITGKETLIDL 170
 QY 208 NTIKNASDSGDGLDLETTNOLKEHPTDFSEVLEFQNSNEVOEFAKAFAYIETQHRDVLQ 267
 Db 171 NKLMLKLYSDEKQLI--EQIYKNNKKEIF--KNIDNVQKEINK-----211
 QY 268 LYAPEAFNYMDKFNEQDINLSLEELKDQRLMSRYEKWEKIKQHYOHWSDSLSEGRGLK 327
 Db 212 -----KDELKNKLDE-----SKKEFIK-----229
 QY 328 KIQIETPEPKKDIIHLSQDEKELKRIODSDSLSTEKEFLKLQIDIRDSLSSEK 387

```

Db      230  -----KQELNMTIDKKQEBELIKKLNDKLEINFNIDEKQKLLDQINKI-NTLNENIK 280
Qy      388  ELLNRIQDVSSNPBLSERKEKEPLKKLKDLPYDINQRLDQTCGLDPSINDVRKQYKR 447
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      281  GYWN-LYETENKKNISLQNEILNK-DSTIKSLDEKQKLLD-----ELDK 322
Qy      448  DIONIDALLHQSLGSLVYKIKYLEENMNNINLTATLGADLVDSNTDKINRGIFNEKKN 507
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      323  NINNTI-----ISLVNK-----SNTKTTNQQLLESSTD-FNNANIN--INELSK 365
Qy      508  FKYSSISNMIYDINERPALDNERLKWRIQSDPTFRAGYLENGKLILQNLGIEIKDYOI 567
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      366  IK-----LPD-NDIOKLNDNITEQNNKIT-----DFNNSTRLFKLEKLDTEYKKIDD 411
Qy      568  IKQSEKEYIRIDAKVYPRKSIDTKIQEAOLNINQENNRKALGLPKTKLTITVYVNR----- 623
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      412  IKNNMLQKLEESYK-----KIDQTEYKKYKKNKINKEINDIETL-KNNNLQKLEENKIDE 465
Qy      624  ---YASNIYESAYLLINEMK-NNIQ-----SDLIKKVYNYLVDCGRFVFDITLP 670
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      466  QTEYKKKNKINKEYNNDIETLKNNNLQKLEENKINDKLTKLNDLESNTFL-----NKL 520
Qy      671  NIAEQYTHODEI-----YEQVHSGGLVYPRSRILHGPSKGVELRNDSECFIHEFGHA 724
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      521  NISDKRKSREIAKLTNTEEQL-----RKDLLENINKTNELMKLSDKRLSSLQOL 570
Qy      725  VDDVAGYL-LDKNOSDLVTNKKFKIDIFKEGSLNLTSGRN--EAFFAFAFLMST 780
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      571  YDSKNILDGIDITKYNLSKEKNKIDETFL--SNIKEPDIVYENKIEIGLNDLSINTKI 626
Qy      781  DHAERLKYQKNAPKTFQFINQIKFLINS 809
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      627  IN-----NDQFKKEYINS 638

RESULT      9
A42771
reticulocyte-binding protein 1 - Plasmodium vivax
C:Species: Plasmodium vivax
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 28-Apr-1995
C:Accession: A42771
R:Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
Cell 69, 1213-1216, 1992
A:Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.
A:Reference number: A42771; MUID:9231538
A:Accession: A42771
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2829 <GAL>
A:Experimental source: Belem strain, merozoites
A>Note: sequence extracted from NCBI backbone (NCBI:108114, NCBI:108115)

Query Match      5.8%; Score 241; Db 2; Length 2829;
Best Local Similarity 18.9%; Pred. No. 0.05;
Matches 194; Conservative 161; Mismatches 326; Indels 346; Gaps 46;

Qy      39  DVGMIYKREKKKKDKNR-----KDEERNKQOEHLKEIMKHIVKI----- 79
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      706  DINALIEVEKVTNKESTLEMLKDEE-----MEKRLDAKETFAKLNFVSDKLTDTYTT 761
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      80  ---EVKGEAVYKKAERLLEKVPDYLEMYKAIGRI-----YIVGD-ITRHI 125
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      762  KMSAEVTNAEGIKKELIAQKQFENVHKKMKFSDARSTFEALQNSMOQYNQGDALIEKHK 821
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      126  SLEALSDEK-----KKIKDIYGDALLHEHYVAKGKY-----EPVLVQSSDDIYENTEK 176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      822  QNRSEKEEYEFKNESVEDLSRETFEEDQYTKHKKNFSRRKQELSAETITMNEVYNKIES 881
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      177  ALNYYEYI-----GKILSRILSKINDPYQKFDVLTMTIKNA--- 213
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      882  QLNATGYLEKTFSLIGDQNEVSTAKALKEKIVSDLSRQKIDQYETEFKEKTSAYENVYST 941

```



```

Db 544 LDKSEENARSIIEVLTCK--EKQKILLENKCNUL--KKQJENKSKNIEE-LHQENKALAKK 599
Qy 462 STLYNK-ITYLENMNNINLTATGADLVDSDTNTKIN-RGIEENKKNFKY-SISSNMI 518
Db 600 SSAENKQOLNAYE-IKVKKLELELA-----STKQKFEEMINNYQIEIETIKKISEEKL 650
Qy 519 VDIENERPALDNERLKMRIQIUSPPTRAGY-----LENGKILIORNIGLEIKVOIIO 570
Db 651 GEVEKAKATVDEAVK--LQKEIDLRCQHKIAEVALMEKHKHQYDKIVESRSESLGYKN 708
Qy 571 SEKEYIHDKAVPKPSKIDTKIOE-----AQJNINQENKALGL-PKTKLTITVNHRY 624
Db 709 REOE--OSSAKVALETELSTNIRELIVLSKLOKEVEKEKEKLMQENTAITIDKKDKI 766
Qy 625 ASNIVEASVILLEMK---NNIOSDLIKVTVLYVDGNC-----RYVFDITLPNIAE 674
Db 767 QASLLESPE--ATSMKFSKTTPPSONISRLSSMSDGSKDNRDSLRAKSAISLTVTYK 824
Qy 675 QYTHODIYEQVHSGKGLYVESRILHPSK-----GVELRNSE-----GTHIFEG 722
Db 825 EYT---VTPPKTSIYOREKNYLTPTGGSNNKRKRVTFEFDVNSQSETTDLISVSE-- 877
Qy 723 HAVDYVAGYLLDNQSD--LYTNSKKEFIDIPKBECSNLSVGRTEAEFFAEPFALMST 780
Db 878 ---EDISNRILYNNNTPDSHLVYTPKQDPLSLSTPASFYKFSLSLKKMR--EDRWATTIAKI 932
Qy 781 DHAERLTQVKNAPKTF 796
Db 933 DRKRRL---KEAEKLF 945

```

RESULT 13
S49461
synaptonemal complex protein 1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-Nov-1999
C:Accession: S49461; S59599
R:Julien, S.; Luc, M.; Francois, C.
A:Submitted to the EMBL Data Library, October 1994
A:Description: Cloning and sequencing of the murine SCP1 cDNA.
A:Reference number: S49461
A:Accession: S49461
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-993 <JUL>
A:Cross-references: EMBL:238118; NID:g1360015; PIDN:CAA86262.1; PID:g55860303
R:Sage, J.; Martin, L.; Cuzin, F.; Rassoulzadegan, M.
Biochim. Biophys. Acta 1263, 258-260, 1995
A:Title: cDNA sequence of the murine synaptonemal complex protein 1 (SCP1).
A:Reference number: S59599; MUID:96004899
A:Accession: S59599
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-993 <SAG>
A:Cross-references: EMBL:238118; NID:g1360015; PIDN:CAA86262.1; PID:g55860303

Query Match	5.5%	Score 230	DB 2	Length 993
Best Local Similarity	21.8%	Pred. No. 0.036		
Matches 202	Conservative 149	Mismatches 346	Indels 228	Gaps 50
QY	44	VKEKKNKDEKRRKDEENKTOEHLKEIMKHIVKEVGEBAVK-----	88	
		: : : : : : :		
Db	124	LKQENKIQENRKIIIEARKA----IQELQFENKVSLSKLEELQDNKDLKENNATIMH	179	
QY	89	---KEAAELLEKVPDVLENYKAIGKIIYI-VDGDIYKH-----SLEALSDEKK---	135	
		: : : : : : : : : : : : : : : : :		
Db	180	CNLLKETCARSAEK--TKMYEYREETQOVYVDLSNIEMKILAEFEELVQENMRLEMH	237	
QY	136	KIKIYQKDALIHENY---YKAGCYEPVLYTQSSQEDVVENTEALNYYETGLISMD	191	
		: : : : : : : : : :		
Db	238	FKLEEDHKIQLHEEYQKEVNNKNQVSELLIOASAK--ENKMDLDFLLE-----ESPD	291	

OY	192	ILSKINOVYOKR-----LDVNTIKNMSDSDGOOL-----LETNOLKCTPTDFSVSEF	238
Db	232	---KANOLEERTKLODENMLKELSEKKDUITSLEDIKMSORMSSTOKALEBDIOIATKT	348
OY	239	LEONANEVOE-----YFKAFAVYIE-----POHNDVLOLAP	271
Db	349	ISO-LTEVKEOMEMELNKAKTHSHVAVIELAKTPTOTLEELRTBEOGRLEKNEDQOLKLTIV	407
OY	272	E-----AFNYWDF-NEOEINLSLEELK-----DQRMLSRYEKWEKIKOHYHWSDSL	319
Db	408	ELOKKSNELEHETKFKNNKEV--ELEELKNIILAEODKILDEKKOYKELAELOEKOEELT	465
OY	320	---EGBRGILKKLOPIERPKKDDIHSLSQE-----EKELKKIQDSS--DPLSTBEK	368
Db	466	FLLETREREVDHLOQVYVTKTSEOHYILKOYEBEMKTELEKEKLTPELTALSCDMLLEAK	525
OY	369	EF-----LKKILODIRDLSSEEEKELNRIQVSSNPLSKE-----KE	407
Db	526	KVVOASDMALELTKKHODIINCKKQOER-LIKOIE-----NLEBKEMHLRELESYKE	579
OY	408	FLK---KLKIDIOPYDINO-----RLDQTGLIDSPSINLDVRYKOYKRDIONIDALL	456
Db	580	FIQOGEVYCKCLDKSEENARSIECEVLKKEKOMLLESKCNUL--KKOVENSKNIEE-L	636
OY	457	HOSIGSTLYNK-----IYLENMNNINLTATIGADLYDSTDNKIRGINFEKKKPK	509
Db	637	HOE-NKTLKXSSAIEKOLNAYE-KVXSLLEL-----ESTKOR--FEBTWNNOQ	683
OY	510	---YSISSNMYIDINERPALDNERLKWIRIOLSPDPRAGY-----LENGKLILOR	556
Db	684	KELEMKKISSEGLLEGEVEKAKATVDEAYK--LOKEILDRCQKIKINEMVALMEKHKHODK	741
OY	557	NIGLEIKDVOIITKOSEKEYIRIDAVVPEKSKIDTIKIOE-----AOJINQEMKNAKGLPK	611
Db	742	IYEEBDSSEIGLYKNNRQOE--OSSAKIALETETELSNIRNELVSLKLOLEIEKEKEKILMAK	799
OY	612	-YTKLITFNVNHRAVSNIYESAYLLINEK--NNIOSLKLKVVNNYVDG-----NG	660
Db	800	ENTALLKDKKKKIDIASLLESPE--ATSKPEKDSKTPPSONISRLSSSDGSKNNRDL	857
OY	661	REVFVDITLPIKAEQYTHODEIYEQVHSH-GLUYEPRSILHGPSKVELNDSE---	715
Db	858	RASAKSILPTIYTKREYLYKTPTKSIIYQRENKIYPTGSGNKKRKAIFEDVMSDSSETAD	917
OY	716	--GFIHEGHAVDVAGYLLDKNOSD--LVYNSKKFIDIFKEEGSNLTASYGTNDAEFPA	771
Db	918	ILLSIYSE-----EDVSNRLYDNNPDSDHILVTKPOTPLSTPSPASFMKFGSLKMYR--E	970
OY	772	EAFRLMHSNDHAERLKYOKNAPKTF	796
Db	971	DMWTIAIKDRKRL--KEAEKLF	992

RESULT 14
T18427
hypothetical protein C0335C - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18427
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18427
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3724 <LAW>
A:Cross-references: EMBL:Z68547, NID:el325376, PID:el325379, PIDN:CA811104.1
C:Genetics:
A:Introns: 307/1, 1545/2
A:Note: C0335C

Query Match	5.58;	Score 229;	DB 2;	Length 3724;
-------------	-------	------------	-------	--------------

Db 1372 VEEGVSGLKKHYDEVW---KYYQKIDKEV-DKEVSKALESKNDVTYVVLKON--QDFFSKY 1425
QY 652 TNTL 655
Db 1426 KNFV 1429

Search completed: December 2, 2001, 13:50:38
Job time: 273 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:46:05 ; Search time 44.8 Seconds

(without alignments)
662.095 Million cell updates/sec

Title: US-09-747-521-2

Perfect score: 4145

Sequence: 1 MNKKEFIKIVSMSCLVTAI.....KNAKPTQFINDIKFTIINS 809

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4145	100.0	809	1	LEF_BACAN
2	504	12.2	800	1	CYAA_BACAN
3	241	5.8	2869	1	RBP1_PLAVB
4	235.5	5.7	2663	1	CENE_HUMAN
5	234	5.6	997	1	SCPL_RAT
6	230	5.5	993	1	SCPL_MOUSE
7	223	5.4	794	1	HMR_MOUSE
8	218.5	5.3	1075	1	Y124_METJA
9	218	5.3	756	1	Y109_YEAST
10	216.5	5.2	1679	1	MYSJ_DICDI
11	215	5.2	2245	1	MYSJ_DICDI
12	212.5	5.1	1957	1	YDB6_SCHPO
13	211.5	5.1	1875	1	MLP1_YEAST
14	211	5.1	1251	1	RBP2_PLAVB
15	210	5.1	1939	1	MYH6_MESAU
16	208.5	5.0	691	1	Y104_YEAST
17	208	5.0	1230	1	SMC3_YEAST
18	208	5.0	2710	1	TOXA_CLODI
19	207	5.0	3911	1	AKA9_HUMAN
20	204	4.9	1312	1	RA50_YEAST
21	204	4.9	1938	1	MYH6_RAT
22	202	4.9	2748	1	NIM1_YEAST
23	201	4.8	1790	1	USO1_YEAST
24	201	4.8	1935	1	MYH7_PIG
25	201	4.8	1940	1	MYH3_RAT
26	200	4.8	845	1	SCPL_MESAU
27	200	4.8	1938	1	MYH6_MOUSE
28	200	4.8	2116	1	MYS2_DICDI
29	198	4.8	1939	1	MYH6_HUMAN
30	197.5	4.8	1427	1	REST_HUMAN
31	197.5	4.8	1805	1	HMW2_MYCGE
32	197.5	4.8	1928	1	MYS1_YEAST
33	196	4.7	976	1	SCP1_HUMAN

34	196	4.7	1939	1	MYH4_HUMAN	Q9Y623 homo sapien
35	194.5	4.7	1727	1	ALM1_SCHPO	Q9UK5 schizosach
36	194.5	4.7	3210	1	CENE_HUMAN	P49454 homo sapien
37	194	4.7	850	1	DPO1_ANATH	Q59156 aneroceallu
38	194	4.7	1940	1	MYH3_CHICK	P02565 gallus gall
39	194	4.7	1978	1	MYH8_CHICK	P10587 gallus gall
40	193	4.7	2136	1	YCF2_MARPO	P09975 marchantia
41	193	4.7	2230	1	GOG4_HUMAN	O13439 homo sapien
42	192.5	4.6	681	1	RPSD_HELPJ	Q9ZMV3 hellicobacte
43	191.5	4.6	944	1	NUF1_YEAST	P32380 saccharomyc
44	191.5	4.6	1164	1	BAG_STRAG	P27951 streptococc
45	190.5	4.6	1818	1	HMW2_MYCPN	P75471 mycoplasma

ALIGNMENTS

RESULT	ID	LEF_BACAN	STANDARD:	PRT:	809 AA.
AC	P15917				
DT	01-APR-1990 (Rel. 14, Created)				
DT	01-APR-1990 (Rel. 14, Last sequence update)				
DT	15-JUL-1998 (Rel. 36, Last annotation update)				
DE	LETHAL FACTOR PRECURSOR (EC 3.4.24.-) (LF).				
GN	LEF.				
OS	Bacillus anthracis.				
OG	Plasmid PX01.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;				
OC	Bacillus/Staphylococcus group; Bacillus.				
OX	NCBI_TaxID=1392;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 34-49.				
RX	MEDLINE=90034185; PubMed=2509294;				
RA	Bragg T.S., Robertson D.L.;				
RT	"Nucleotide sequence and analysis of the lethal factor gene (lef)				
RT	from Bacillus anthracis.";				
RL	(Gene 81:45-54(1989)).				
RP	SEQUENCE FROM N.A.				
RA	Lowe J.;				
RL	Submitted (APR-1990) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	ZINC-BINDING.				
RX	MEDLINE=95154669; PubMed=7851740;				
RT	Kochi S.K., Schiavo G., Mock M., Montecucco C.;				
RL	"Zinc content of the Bacillus anthracis lethal factor.";				
CC	FEMS Microbiol. Lett. 124:343-348(1994).				
CC	- FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN.				
CC	AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE				
CC	DEATH. LE IS THOUGHT TO BE A LETHAL FACTOR THAT, WHEN ASSOCIATED				
CC	WITH PA, CAUSES DEATH. LE IS NOT TOXIC BY ITSELF. PA IS THOUGHT TO				
CC	BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC CELLS, THEREBY				
CC	FACILITATING THE INTERNALIZATION OF LE OR EF.				
CC	- SUBUNIT: SECRETED ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT				
CC	PROTEINS: A PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN				
CC	EDEMA FACTOR (EF). NONE OF THESE IS TOXIC BY ITSELF.				
CC	- SUBCELLULAR LOCATION: SECRETED.				
CC	- SIMILARITY: THE PA-BINDING REGION IS FOUND IN BOTH B. ANTHRACIS EF				
CC	AND LE.				
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M34 (ZINC				
CC	METALLOPROTEASE).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL; M29081; AAA79216.1; -				
DR	EMBL; M30210; AAA22569.1; -				

PIR: J00032. J00032.
DR MEROPS: M34.001.34.
DR InterPro: IPR003541: Anthrax toxinA.
DR InterPro: IPR000130: Zn_MTPeptide.
DR PROSITE: PS00142: ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Toxin; Signal; Repeat; Plasmid.
FT SIGNAL 1 33
FT CHAIN 34 809 LETHAL FACTOR.
FT DOMAIN 34 293 PA-BINDING REGION (POTENTIAL).
FT DOMAIN 300 420 REPEATS.
FT METAL 719 719 ZINC (CATALYTIC) (POTENTIAL).
FT ACT_SITE 720 720 POTENTIAL.
FT METAL 723 723 ZINC (CATALYTIC) (POTENTIAL).
SQ SEQUENCE 809 AA; 93786 MW; 8C1684D7277310AE CRC64;

Query Match 100.0%; Score 4145; DB 1; Length 809;
Best Local Similarity 100.0%; Pred. No. 1.2e-153;
Matches 809; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKKKFKVISMGLVTAITLSCGPVPIPLVQAGAGHDVGMHYKEKKNKDKRDEE 60
DB 1 MNKKKFKVISMGLVTAITLSCGPVPIPLVQAGAGHDVGMHYKEKKNKDKRDEE 60
QY 61 RNKTOEHLKIMKHIVKIEYKGEAVYKKAEEKLEKVPSPDYLEMYKATIGKITYVGD 120
DB 61 RNKTOEHLKIMKHIVKIEYKGEAVYKKAEEKLEKVPSPDYLEMYKATIGKITYVGD 120
QY 121 ITHKISLEASDEKRIKIDYGDALLHEHYVYAKEGEPVLIQSSHDYVENTEKALNV 180
DB 121 ITHKISLEASDEKRIKIDYGDALLHEHYVYAKEGEPVLIQSSHDYVENTEKALNV 180
QY 181 YYEIGKILSRILSKINOPYKFLDVLNTIKNASDSGQDLFTNQLEKHPDTSVEFLE 240
DB 181 YYEIGKILSRILSKINOPYKFLDVLNTIKNASDSGQDLFTNQLEKHPDTSVEFLE 240
QY 241 QNSNEVOEVFAKAFAYYIEPOHRDVLQYAEAFYMYKFMDEQINLSLEEKQRMISR 300
DB 241 QNSNEVOEVFAKAFAYYIEPOHRDVLQYAEAFYMYKFMDEQINLSLEEKQRMISR 300
QY 301 YEKWEKIKOHYOHMSDSLSEEGRLKKLQPIPEPKDDIHSLSOEKELKRIQIDSS 360
DB 301 YEKWEKIKOHYOHMSDSLSEEGRLKKLQPIPEPKDDIHSLSOEKELKRIQIDSS 360
QY 361 DFLSTEKEEFLKQIDIRDSLSEBEKELNRIOVDSNPSEKKEFLKRLKLDIQPYD 420
DB 361 DFLSTEKEEFLKQIDIRDSLSEBEKELNRIOVDSNPSEKKEFLKRLKLDIQPYD 420
QY 421 INORLOPTGGLIDSPSTLDVRYKKRDIONIDALLHOSIGSTLYNKTYLENNMINLT 480
DB 421 INORLOPTGGLIDSPSTLDVRYKKRDIONIDALLHOSIGSTLYNKTYLENNMINLT 480
QY 481 ATTAGADLVSDTNKINNGINFEKKNKYSISSNYMIVDINERPALDNERLKRIQISP 540
DB 481 ATTAGADLVSDTNKINNGINFEKKNKYSISSNYMIVDINERPALDNERLKRIQISP 540
QY 541 DTRAGYLENGLIIQNRINGEIKVQIIKQSEKEYIRIDAKVYKSKIDTKIOEAQLIN 600
DB 541 DTRAGYLENGLIIQNRINGEIKVQIIKQSEKEYIRIDAKVYKSKIDTKIOEAQLIN 600
QY 601 QENKALGLPEYTKLTFENVNRASNTVESAYILNEMKNNISDILKKTATNLVQDNG 660
DB 601 QENKALGLPEYTKLTFENVNRASNTVESAYILNEMKNNISDILKKTATNLVQDNG 660
QY 661 RFVFTDITLPIAEOYTHODEIYEQVSHKGLYVESRSILLHGSPKGYELANDSEGFIHE 720
DB 661 RFVFTDITLPIAEOYTHODEIYEQVSHKGLYVESRSILLHGSPKGYELANDSEGFIHE 720
QY 721 FGAAVDDYAGYLLDKNOSDLVTSNKKFIDIFKEGSGNLTSGRTNEAEFFAEARLMHST 780
DB 721 FGAAVDDYAGYLLDKNOSDLVTSNKKFIDIFKEGSGNLTSGRTNEAEFFAEARLMHST 780
QY 781 DHAERLKVQKNAPKTFQFINQIKFIINS 809

DB 781 DHAERLKVQKNAPKTFQFINQIKFIINS 809
RESULT 2
CYAA_BACAN STANDARD: PRT: 800 AA.
ID CYAA_BACAN
AC P40136;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CALMODULIN-SENSITIVE ADENYLATE CYCLASE PRECURSOR (EC 4.6.1.1) (ATP
DE PYROPHOSPHATE-LIASE) (ADENYLYL CYCLASE) (EDEMA FACTOR) (EF).
GN CVA.
OS Bacillus anthracis.
OG Plasmid PX01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
CX NCBI_TaxID=1392;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=89138004; PubMed=2906312;
RA Escuyer V., Duflot E., Sezer O., Danchin A., Mock M.;
RT "Structural homology between virulence-associated bacterial adenylate
RT cyclases.";
RL Gene 71:293-298(1988).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=89211974; PubMed=3149607;
RA Robertson D.L., Tippetts M.T., Leppla S.H.;
RT "Nucleotide sequence of the Bacillus anthracis edema factor gene
RL (cya) a calmodulin-dependent adenylate cyclase.";
RN [3]
RN SEQUENCE FROM N.A.
RA Escuyer V., Duflot E., Mock M., Danchin A.;
RT "Nucleotide sequences expressing adenylate cyclase from B.anthraxis,
RT proteins having the activity of this adenylate cyclase and biological
RT uses.";
RL Patent number EP0366550, 02-MAY-1990.
RN [4]
RN SEQUENCE FROM N.A.
RX MEDLINE=88198021; PubMed=2834337;
RA Tippetts M.T., Robertson D.L.;
RT "Molecular cloning and expression of the Bacillus anthracis edema
RT factor toxin gene: a calmodulin-dependent adenylate cyclase.";
RL J. Bacteriol. 170:2263-2266(1988).
RN [5]
RN SEQUENCE OF 34-48.
RX MEDLINE=89211974; PubMed=3149607;
RA Schmidt J.;
RL unpublished results, cited by:
RL Robertson D.L., Tippetts M.T., Leppla S.H.;
RL Gene 73:363-371(1988).
RN [6]
RN REVIEW.
RX MEDLINE=93119764; PubMed=8418825;
RA Danchin A.;
RT "Physiology of adenylate cyclases.";
RT Adv. Second Messenger Phosphoprotein Res. 27:109-162(1993).
CC -1- FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,
CC AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE
CC DEATH. EF IS A CALMODULIN-DEPENDENT ADENYLYL CYCLASE THAT, WHEN
CC ASSOCIATED WITH PA, CAUSES EDEMA. EF IS NOT TOXIC BY ITSELF. PA IS
CC THOUGHT TO BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC CELLS,
CC THEREBY FACILITATING THE INTERNALIZATION OF EF OR LF.
CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC -1- ENZYME REGULATION: ACTIVATED BY HOST CALMODULIN.
CC -1- SUBUNIT: ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT PROTEINS, A
CC PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN EDEMA FACTOR
CC (EF). NONE OF THESE IS TOXIC BY ITSELF.
CC -1- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-2 FAMILY.
CC -1- SIMILARITY: THE PA-BINDING REGION IS FOUND IN BOTH B.ANTHRACIS EF

CC AND LF.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M23179; AAA22374.1; -
 DR EMBL: M24074; AAA9215.1; -
 DR EMBL: A07289; CAA00652.1; ALT_SEQ.
 DR Interpro: IPR003541; Anthrax_toxinA.
 KW Lyase: CAMP synthetase; Toxin; ATP-binding; Signal; Plasmid.
 FT SIGNAL 1 33
 FT CHAIN 1 33
 FT DOMAIN 34 800 CALMODULIN-SENSITIVE ADENYLATE CYCLASE.
 FT NP_BIND 289 680 PA-BINDING REGION (POTENTIAL).
 FT NP_BIND 347 354 CATALYTIC.
 FT CONFLICT 350 350 V -> E (IN REF. 2).
 FT CONFLICT 510 510 Q -> T (IN REF. 2).
 FT CONFLICT 512 513 EM -> RM (IN REF. 2).
 FT CONFLICT 760 760 V -> L (IN REF. 3).
 FT SEQUENCE 800 AA; 92477 MW; F4F7EB485DF4C5A6 CRC64;
 Query Match 12.2%; Score 504; DB 1; Length 800;
 Best Local Similarity 24.1%; Pred. No. 4.9e-13;
 Matches 214; Conservative 154; Mismatches 297; Indels 224; Gaps 40;
 Oy 4 KKEFT---KVISMGLVATITLGPVFIIVQAGAGHGDVGMHYKREKKNDENRKDE 59
 Db 3 RKKEFTPKNFSTISFVLLPAISSQAIEVNA-----MHEHTESDIKNNHT 49
 Oy 60 ERNKTOEHLKEIMKHIVKIEVGEAVKKEAEKLEKPEPSDVEIMYKAIGKIYVDG 119
 Db 50 EKNKTEKEKFKDSINNVLVTEFTETLTKIQOTDLKKIPROVLEIYSELGGEIYFTDI 109
 Oy 120 DITKITSLEALSEDKKIKIDYIGKDALHEHYVAKEGEPVLYVIOSSDYVENEKALN 179
 Db 110 DLVEHKELODSEEEKNSNSRGEVFPASRFVEKKRTPLLII-NIDVYAINSSOSKE 168
 Oy 180 VYEELGKILSPILSKINQPYOKPLD-VLNTIKNAS-DSDODILFTNOLAE---HPT 232
 Db 169 VYEEIGKGISDIISK-----DKSIDPEFLNLKISLSDSDSDDLFSCKFEKLELNK 223
 Oy 233 DESVEFLQNSNEVOEVPKAFAYYIEPOHRDVLQIYAPAEFNYMDKFNEDINLSLEEL 292
 Db 224 SIDINFIKENLTFQHAFSLAFSYFAPDRHRTVLELYAFDMFEYNNKL----- 271
 Oy 293 KDQRLSLREK-WEKIKOHYHMSDSISEGRGLIKLQIDIEPKKDDIHSLSOEKE 350
 Db 272 -----EKGGFEKI-----SESLKKEG-----VEKDRIDVL---KGEKA 301
 Oy 351 LKRIQIDSDPLSEKEKFKLQIDIRDSISEEKELNRIQVDSNPLSEKEKFEFLK 410
 Db 302 L-----KASGLVPEHDAFKKI-----ARELNTYILFEPVKLNTN---LI 339
 Oy 411 KLIKIDQPYDINQRIQDGLIDSPSINDVKKOY-----KRDIONIDALL-HQ-SIG 461
 Db 340 KSGVATKGLNIVKSGSSDMQPVAGIYIPFODDLSKKHQOOLAVEKGNLENKKSITHEHGELG 399
 Oy 462 STLYNKKIYL-YENNNINNLATLGLADLVSDNTKINRGIFNEFKNFK-----YSISSN 515
 Db 400 -----KIPKLKDLHRIIEELK-----ENGIIILKKEKKEINDGKKYVLLSEN 438
 Oy 516 YVIIVDINERPALDNERLKWRIQISPDTRAGYLE-NCKLLQNLNIGLEIDVO-IIKOSKE 573
 Db 439 NOYIE--RISDENNEVOYK-----TKGKITVLEGEKFNMRNIEYMANVGEVLKPLTA 490
 Oy 574 EYIRIDAKVVPKSIDTKIOEQALNINQEMNKALGLPKYTK-----LITFNVNRYA 625

Db 491 DY---DLFALAPSLTEIRKOIPQ-----KEWDKVVNTPNSLEKQKVYLLIKYIERK-- 541
 Oy 626 SNVIESAYILNKNKNNIOSDLIKKTYNLYVONGNRFVFDITLPLNIADYTHOD--EIV 663
 Db 542 ---PDSYKCTLSWQKOMDRL-----NEAVKTYGYGGVYVNHGTBDNEDEFP 567
 Oy 684 EGVHSGLYVPEERSILLHG---PSKGVELRNDSEGFIEFGVADVDA-----GYL---- 732
 Db 588 EKNELFEIINPEEFLITKNWEMTGRIEKNITGKQIYLYFNNSYKKAIAQNKAYLEMD 647
 Oy 733 -LDKNOSDLVTNSKFTI-----DIEKEGNSN-----LTSYGR 764
 Db 648 PIYKAKINPTSAEFIKNLSSIRSSNNGVYKSDKDEFAKESVKKIAGYLSDYNS 707
 Oy 765 NEKEFFAE-----AFRLHSTDAER-LKVQNAK---TFPINDOI 803
 Db 708 ANHIFOEKRRKISIRGIAVMEIENVLKSKOIAEYKNFYOLKERI 756
 RESULT 3
 ID RBPI_PLAVB STANDARD; PRT; 2869 AA.
 AC Q00798;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
 GN RBPI.
 OS Plasmodium vivax (strain Belem).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=31273;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92315338; PubMed=1617731;
 RA Galinski M.R., Medina C.C., Ingavallio P., Barnwell J.W.;
 RT "A reticulocyte-binding protein complex of Plasmodium vivax
 RT mercoites";
 RL Cell 69:1213-1226(1992).
 CC -I- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
 CC HUMAN RETICULOCYTE CELLS.
 CC -I- SUBUNIT: HOMODIMER (POTENTIAL).
 CC -I- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M88097; AAA29743.1; -
 DR HSSP: P36956; IAM9.
 KW Malaria; Receptor; Signal; Transmembrane.
 FT SIGNAL 1 17
 FT CHAIN 1 17
 FT DOMAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
 FT TRANSMEM 2808 2826 EXTRACELLULAR.
 FT DOMAIN 2827 2869 POTENTIAL.
 FT SITE 1030 1032 CYTOPLASMIC.
 FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
 FT SEQUENCE 2869 AA; 330213 MW; B9DBE442205BECFF CRC64;
 Query Match 5.8%; Score 241; DB 1; Length 2869;
 Best Local Similarity 18.9%; Pred. No. 0.032;
 Matches 194; Conservative 161; Mismatches 326; Indels 346; Gaps 46;
 Oy 39 DVGMYKREKKNDENRK-----KDEERNKTOEHLKEIMKHIVK----- 79
 Db 746 DIALIEVEKEFVTENKKESTLEMLKDEE---MEKLODAKFAKLNFEVSDKLTVDYVT 801
 Oy 80 ---EVKGEAVKKEAEKLEKPEPSDVEIMYKAIGKI-----YIVDD-ITKIH 125


```

OY 82 KGEAVKKEAEKLEKVPDVLNEMKAIGKITYVDGDTFKHLSLESDKKIK--- 138
DB 584 LRE-----KEDQIKKLEIDSKLENIMK-----DLS--YISLEST-EDPKOMKOL 627
OY 139 -----DIYKDALLHEHYVAEGEYEPVLVIOSSSEDEVVENTEKALNVYEIGKILSR 190
DB 628 FDAETVALDAKRESAFILSENLELEKEMKE---LATYKQEMENDIOLYOSQLEAKKKMVO 684
OY 191 DILSLINOPYOKFLDVLNLIKASDSG---ODLFTNOLKEHPDPSVEF---LEQNSN 244
DB 665 DLEKELQSAFNITRLTSLI-----DGKVPDLCNLLECKITDLOKELKEVEENEA 738
OY 245 EKOGEF-----AKAFAYTIEPOHRD-----LOLAPFAFNMDKF-----NEQENLNS 288
DB 739 LREYVILSELKSLPSEVRLKKEIQDKSEELHITSE---KDLFSEYVAKESVQGL 794
OY 289 LEEL---KDQRLMSRYEKMEKIKOHYQWMSDSLSEGRGLKLLOIPIBKDDIHS-- 343
DB 795 LEETGKTDD-LATQSNYKSTDOEFQNF-----KTLHMDFOQTKMYLEENE 841
OY 344 -LSQEKELKRIQ-IDSS-DPLSTE-----EKFLKLQIDTD 380
DB 842 RMOQEVINLSKEAQKDFSSGLAKTELSTYQLOEKTREVOERLNMEDQKE-QLEEND 900
OY 381 S---LSEEEKELNRIQVDSNPLSEKEKFLKLIDIPYDINQRLQDT----- 428
DB 901 SPLQVVERKTL-----ITEKLOQTLSEVKVTLQOEKDLQKQDSIQIENDQK 949
OY 429 GGLIDSPSINDVROKQKRDIONIDALLHQSIGSTLYNKIYENNNINLTATIGADLV 488
DB 950 SGIHDTVMNMIIDTQOLRMALESLEKQ--HQETINTLKSISEEVSRLH-----ME 998
OY 489 DSTDMTKINRGIFNEPKKFKYSISSNMYIVDINERPALDNERLKRLOLSPDTRAGIYE 548
DB 999 EMTGETK-----DEFQOK-----MVGIDKKQDLEANN----- 1025
OY 549 NGKLLIQRNIGLEIKDVQITIKQSEKERYIRIDAKVVPKSIDTKIOE-AQLINQEWNNKAL 607
DB 1026 -----TOTLTDADVNDNEITEBOQRKIFSLQEKNELOQMLESVIAKEBOLKTDKKN--- 1076
OY 608 GLPKYTKLTFFNVNHRKASINIVESAVLLINENKNNIQSLIKVTVYLVGNGRFVETDI 667
DB 1077 -----IEMTIEHQ-----EELRLLDGLDELK--QOEIVAOEKHNAIKKEGELSRTCD 1120
OY 668 TLPNIAEOYTHODEIYEOVHSGKLYVPERSJILLHGPSKVELRNDSEGFIEHFGAVND 727
DB 1121 RLAEVEEKLEKESQOLQOEKQOOLLNVQEMSEMOKKINIEIMKLELKRKELTLEHMETE 1180
OY 728 --YAGILDKNOSD--VTNSKRFI---DIFKEEGSNLTSYGRNVEAEFFAEAFRL--- 776
DB 1181 RELQAKLKNENEEVKSTIKERKVKLELOKSFETTERDHLRGITREIYAGLOTKKEELKIA 1240
OY 777 -MHSFDAERLKVQKN-APKTFQFIINDQ 802
DB 1241 HILKEHOETIDELRVSSEKTAQIINTQ 1269

RESULT 5
SCPI_RAT STANDARD: PRT: 997 AA.
AC Q03410:
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SYNAPTONEMAL COMPLEX PROTEIN 1 (SCP-1 PROTEIN).
GN SCPI1 OR SCPI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=Testis;
RX MEDLINE=93099884; PubMed=1464329;
RA Meuwissen R.L.J., Offenberg H.H., Dietrich A.J., Riesewijk A.,
RA Terseel M., Heyting C.,
RT "A coiled-coil related protein specific for synapsed regions of
RT meiotic prophase chromosomes.",
RL EMBO J. 11:5091-5100(1992).
CC -1- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF
CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS
CC CHROMOSOMES DURING MEIOTIC PROPHASE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF
CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.
CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS
CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE
CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE
CC SYNAPTONEMAL COMPLEX.
CC -1- TISSUE SPECIFICITY: TESTIS.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EXCLUSIVELY IN MEIOTIC PROPHASE
CC CELLS.
CC -1- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,
CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL
CC DOMAIN HAS DNA-BINDING CAPACITY.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION, A FRAMESHIFT WAS
CC CORRECTED IN POSITION 6 TO MAXIMIZE THE SIMILARITY WITH THE
CC OTHER SPECIES SCPI1 SEQUENCES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: X67805; CAA8006.1; ALT_FRAME.
DR PIR: S28061; S28061.
KW Nuclear protein; Meiosis; Cell division; Phosphorylation;
KW DNA-binding; Coiled coil.
FT DOMAIN 108 819 COILED COIL (POTENTIAL).
FT DOMAIN 118 121 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 701 704 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 902 905 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 982 990 ARG/LYS-RICH (BASIC).
SQ SEQUENCE 997 AA; 116511 MW; 229D59823FD684BE CRC64;

Query Match 5.6%; Score 234; DB 1; Length 997;
Best local Similarity 21.7%; Pred. No. 0.017;
Matches 199; Conservative 158; Mismatches 349; Indels 210; Gaps 50;

OY 44 VKKEKNKDNKKRDKDERNKTOEHLKEIMKHLVKE-----VKGEAVK----- 88
DB 128 LKQENKLOENRKLIEQKRAIOELOFENKESVLKLEEIQENKDKLKENNATRHMCNLL 187
OY 89 KEAEKLEKLEKVPDVLNEMKAIGKITYI-VDGDITKHI-----SLFALSDEKK-----KIK 138
DB 188 KETCARSAEK--TSKYEVEEETROYVVDLNNIEMKILAFELRYQAEARLEMEFKIK 245
OY 139 DIYKDALLHEHY-----VYAKGEYEPVLVIOSSSEDEVVENTEKALNVYEIGKILSRDILSK 195
DB 246 EDEKIQHLEEEYQKEVNNENOVSLILIOSTER--ENKMKDLTFLE-----ESRD--K 296
OY 196 INOPYOKF-----LDVLNLIKASDSGODLFTNQ-----LKEH-----PTDSVEF 238
DB 297 ANOLEKTKIQDENLKELEKDKHLTSELBDIMWSQKRSSTOKTLEEDIQATYTKIYQL 356
OY 239 LEQNSNEVOEV-----FAKAFAYTIE-----POHRDVLQOLYAPE----- 272
DB 357 TEKEKQOMELNKAKTTHSLVYELKATCTLEELRTQEQRLNEDDQAKLTLMELQKK 416
OY 273 --AFNYMDKF-NEQELNLSLEELK-----DQRLMSRYEKMEKIKOHYQWMSDSL---EE 321
DB 417 SSELKMTKFKNNKFEV--ELEELKTLIAEDQKLLDEKKQVEKLAELQGEKQDLTFLQOT 474

```

```

OY 322 GRGLKKQIDPIEKKODIHSLSOE-----EKELIKRIQID-SSDFLSTEEKF--- 370
D 475 REKEIHLEVOVNTKSEEHYLKQVEEMKTELKEKLIKNIELANSMDLLENKKLVQ 534
OY 371 -----LKKQIDIRSLSEEEK-----ELLNIOVDSSNPSEKEKEFLK-----KTK 413
D 535 ASDVAVLELKHQEDIIINCKQOEEMKQIETLEKEKMLRDELSVREKFTQGDDEVCK 594
OY 414 LDIO-----PYDINORLQDTGLIDPSINLDPRKQYKRDIONIDALHQ-----SIG 461
D 595 LDKSEENARSIEYEVULK-EKOMILEKKNL--KKQIEKSKNIEE-LHQENKALKKK 650
OY 462 STLYNK-IYLYENANNINLTITLADLVSDTNTKIN-RGIFNEKKNFY-SISSNMI 518
D 651 SSANCKOLNAE-IKVKLELELA-----STKOFEEEMINNYOKEIEIKISEEKL 701
OY 519 VDIENERALDERLKWIRIOTSPDRAGY-----LENGLLIQINGLIKVOYIIO 570
D 702 GEVKKAKATYDEAVK--LQKEIDLRQCKIAEMVALMEKHQYDKIVEERDSEGLYKN 759
OY 571 SEKEYIRIDAKVVKSKIDTRKIQE-----AQLNIOENKALG-PYTKLITFNVHRY 624
D 760 REOE--QSAKVALETETLSNIRNELVSLKKQLEVEKEKEKLMQENTALTITKCKK 817
OY 625 ASNIVESAYILINEMK--NNIOSDLIKVTNYIVDNG-----RPFYDTITLPNIAE 674
D 818 QASLLESPE--ATSMKFDSTKTPSQNISRLSSMDSGSKDNDRSLASAKSILSTYTK 875
OY 675 QYTHQDIEYEQVHSGLYVPSRSILHGPSK-----GVELRANDE-----GFINEFG 722
D 876 EYT-----VTPPKTSYIQRBNKTLPPGSGSKKRYTFEPFVNSDSETTDLISLVE-- 928
OY 723 HAVDYAGYLLDKNQSD--LVTNKSKETIDIFKEEGSNLTSYGTNEAEFEAFELMHST 780
D 929 ---EDISNRIYNNNTPSHLLVTKPKQPLSLSTPASTTKGSLKMR--EDRWATIAKI 983
OY 781 DHAERLKVOKNAPKTF 796
D 984 DRKRL--KEAEKLF 996

```

```

RN [4]
RP SEQUENCE OF 95-787 FROM N.A.
RC STRAIN=ICR; TISSUE=Testis;
RA Tsuchida J., Nishina Y., Nozaki M., Uchida K., Nishimune Y.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF
CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS
CC CHROMOSOMES DURING MEIOTIC PROPHASE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF
CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.
CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS
CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE
CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE
CC SYNAPTONEMAL COMPLEX (BY SIMILARITY).
CC -1- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,
CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL
CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z38118; CAAB6262.1; -
DR EMBL: LA1069; AAA64514.1; ALT_INT.
DR EMBL: U62864; AAC53335.1; -
DR EMBL: U62860; AAC53335.1; JOINED.
DR EMBL: U62861; AAC53335.1; JOINED.
DR EMBL: U62862; AAC53335.1; JOINED.
DR EMBL: U62863; AAC53335.1; JOINED.
DR EMBL: D88539; BAAL3639.1; -
DR TMD: MGI:105931; Sycp1.
KW Nuclear protein; Meiosis; Cell division; Phosphorylation;
KW DNA-binding; Coiled coil.
FT DOMAIN 12 97
FT DOMAIN 104 815 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 114 117 COILED COIL (POTENTIAL).
FT DOMAIN 697 700 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 898 901 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 978 986 ARG/LYS-RICH (BASIC).
FT CONFLICT 527 527 F -> L (IN REF. 2).
SQ SEQUENCE 993 AA; 115962 MW; 1AAFA790D64FAFE6 CRC64;

```

Query Match 5.58; Score 230; DB 1; Length 993;
 Best Local Similarity 21.88; Pred. No. 0.024;
 Matches 202; Conservative 149; Mismatches 346; Indels 228; Gaps 50;

```

OY 44 VKEKKKNKDEKRRKDEERNKTOEHLKEIMKHVYKIEVKGGEAVK----- 88
D 124 LKQENKLNQEKRIITEQKRA-----IQELOENKVSILKEEIQENKDLKENNATIIHW 179
OY 89 -----KEAEKLEKVPDVLVEMKAIQKIYI-VDGDITKHI-----SLAESDEKK--- 135
D 180 CNLKEKFCARSAEK--TNKYEYEREETROYVDLNSNIEKKIILFAEELRVOAEKARLEMH 237
OY 136 -KIIDYIGKALLHEHY---VYAKEGYEPVLIYDSSSEDIYVENTKALNVIYEIKIISRD 191
D 238 FKLEDEHKIQLHEEYQKEVKNKENVSELLIQAER--ENKKKDLTFPLE-----BSRD 291
OY 192 ILKINQPYQKF-----LDVLNTRKNASDSDGDL-----LFTNQLKEHPDPSVEF 238
D 292 ---KANOLEETKTIQDENLKEKSKDHLTSELEDIKMSQKRSSTOKALEEDIQITKT 348
OY 239 LEQNSNEVQE-----VFAKAFAYIE-----POHRDVLQDYAP 271
D 349 ISQ-LTEVKEAQEMELKNKAKTTTSHFVYTELKATCTLEELLRTPEQRLKKNEDQKLITV 407
OY 272 E-----AFNVMKGF-NEQELNLSLEELK-----DQIMLSYEKWEKIKQHYQHWSDSL 319

```

```

Db 408 ELQKSNLEEMTKFNKNKV--ELEELKNILAEQDKLDEKQVKEAELOEKDEQELT 465
Oy 320 ---EEGRGLKKQIPIEPKDDIHSLSOE-----EKELKRIQIDSS--DFLSTEEK 368
Db 466 FLLETRKEKVDLOEYATYTKSEQHYLKQVEEMKTELEKEIKKNTLTAACMLLENK 555
Oy 369 EF-----LKLQIDIRDSLSEEEKELNRIQVDSNPLSEKE-----KE 407
Db 526 KFOVASDMALFKKHQEDINCKQOER-LKQIE-----NLEEKEMHLRDELSVRE 579
Oy 408 FLK-----KLKDIQYDINO-----RLQDTGGLDSISINDVKQYKRIQNTDALL 456
Db 580 FIOQGEVCKLQDKSEENRSIECEVLKKEQKILKSCNNL--KQOYENKSKNTEE-L 636
Oy 457 HOSIGSTLKNK-----LYLENNMINULTAGADVSTQNTKINGIFENFKNFK 509
Db 637 HOB-KNTLKKSSAEIKQLMAYE-IVSKLELEL-----ESTQR---FEEMTNNTQ 663
Oy 510 -----YSSSNMYVIDINERPALDNERLKWRIQISPDTRAGY-----LENGKLILOR 556
Db 684 KEIENKKISEGKLGEVERAKATVDEAVK--LQKEIDLQCHRIAMVALMEKHKKHQYDK 741
Oy 557 NIGLEKNDVOIIKQSEKEIRIDAKVPRKSIDTKIQE-----AQUNIDENKKAIGLPR 611
Db 742 IVEERDSEGLYKKNROE--QSSAKTALLETLSNINRELVSILKQLEIEKEEKLKMAK 799
Oy 612 -YTKLITFVNHRYASNIYESAVLLENK---NNIQSLIKKVTYVLDG-----NG 660
Db 800 ENTALILDKDKKIQASLESEP--ATSWKFPDSKTPPSONISILSSSMQSGSKDNRL 857
Oy 661 REVFTDITLPNIAEOYTHODEIYEQVHSK-GLYVPESSRILLHGPKGVELRNDSE---- 715
Db 858 RASAKSILFTYTKETVTPTKKSIYQRENKTIPIGSGNKKRKTAFEDVNDSSSETAD 917
Oy 716 -GFHEFGHAVDYGAYLLDNQSD--LVYNSKFIIDFKKEGSLUTJSGRTNEAEFFA 771
Db 918 LLSLVSSE-----EDVSNRLYDNNPPDSHLVTKPTPLSLTPSPAFMKGSLKKMR--E 970
Oy 772 EAPRLMSTDHMERLKVKONAPKTF 796
Db 971 DKMTTIAKIDRRRL--KEAEKLF 992

```

```

RX MEDLINE-99107769; PubMed-9889313;
RA Fieber C., Plug R., Sleeman J., Dall P., Ponta H., Hofmann M.;
RT "Characterization of the murine gene encoding the intracellular
RL hyaluronan receptor IHABP."
RN Gene 226:41-50(1999).
[4]
RP SEQUENCE OF 164-794 FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN-BALB/C; TISSUE-Fibroblast;
RX MEDLINE-96011639; PubMed-7590272;
RA Entwistle J., Zhang S., Yang B., Wong C., Li Q., Hall C.L., A.J.,
RT Kowat M., Greenberg A.H., Turley E.A.;
RL "Characterization of the murine gene encoding the hyaluronan receptor
RH RHAMM."
RN Gene 163:233-238(1995).
[5]
RP SEQUENCE OF 318-794 FROM N.A., AND CHARACTERIZATION.
RC STRAIN-BALB/C;
RX MEDLINE-92299690; PubMed-1376732;
RA Hardwick C., Hoare K., Owens R., Hohn H.P., Hook M., Moore D.,
RT Criggs V., Austen L., Nance D.M., Turley E.A.;
RL "Molecular cloning of a novel hyaluronan receptor that mediates tumor
RT cell motility."
RL J. Cell Biol. 117:1343-1350(1992).
RN [6]
RP CHARACTERIZATION.
RX MEDLINE-94308286; PubMed-7518470;
RA Hall C.L., Wang C., Lange L.A., Turley E.A.;
RT "Hyaluronan and the hyaluronan receptor RHAMM promote focal adhesion
RL turnover and transient tyrosine kinase activity."
RN J. Cell Biol. 126:575-588(1994).
[7]
RP ERK REGULATION, AND SUBCELLULAR LOCATION.
RX MEDLINE-98225222; PubMed-9556628;
RA Zhang S., Chang M.C., Zylka D., Turley S., Harrison R., Turley E.A.;
RT "The hyaluronan receptor RHAMM regulates extracellular-regulated
RL kinase."
RN J. Biol. Chem. 273:11342-11348(1998).
[8]
RP REVIEW.
RX MEDLINE-99059494; PubMed-9845361;
RA Hofmann M., Assmann V., Fieber C., Sleeman J.P., Moll J., Ponta H.,
RT Hart I.R., Herrlich P.;
RL "Problems with RHAMM: a new link between surface adhesion and
RT oncogenesis?"
RN Cell 95:591-592(1998).
[9]
RP -1- FUNCTION: INVOLVED IN CELL MOTILITY. WHEN HYALURONAN BINDS TO
CC HMHR, THE PHOSPHORYLATION OF A NUMBER OF PROTEINS, INCLUDING THE
CC FOCAL ADHESION KINASE OCCURS. MAY ALSO BE INVOLVED IN CELLULAR
CC TRANSFORMATION AND METASTASIS FORMATION, AND IN REGULATING
CC EXTRACELLULAR-REGULATED KINASE (ERK) ACTIVITY.
CC -1- SUBUNIT: SUBUNIT OF THE HARC COMPLEX.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE AND CYTOPLASMIC.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; RHAMM1V4 (SHOWN HERE) AND
CC RHAMM1; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBIOUITOUSLY EXPRESSED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF031932; AAC12655.1; -
DR EMBL: AF079222; AAD08670.1; -
DR EMBL: AJ005919; CAA06768.1; -
DR EMBL: AJ005920; CAA06768.1; JOINED.
DR EMBL: AJ005921; CAA06768.1; JOINED.
DR EMBL: AJ005922; CAA06768.1; JOINED.
DR EMBL: AJ005923; CAA06768.1; JOINED.
DR EMBL: AJ005924; CAA06768.1; JOINED.
DR EMBL: X64550; CAA5849.1; -

```

DR EMBL: X64550; CAA45848.1; -
 MGD: MGI:104667; Homr.
 KM Hyaluronic acid; Alternative splicing; Repeat; Glycoprotein.
 FT DOMAIN 442 546 5 x 21 AA NEAR PERFECT TANDEM REPEATS.
 FT REPEAT 442 462 1.
 FT REPEAT 463 483 2.
 FT REPEAT 484 504 3.
 FT REPEAT 505 525 4.
 FT REPEAT 526 546 5.
 FT DOMAIN 719 729
 FT DOMAIN 741 750
 FT CARBOHYD 53 53
 FT CARBOHYD 134 134
 FT CARBOHYD 279 279
 FT CARBOHYD 446 446
 FT CARBOHYD 467 467
 FT CARBOHYD 488 488
 FT CARBOHYD 509 509
 FT CARBOHYD 530 530
 FT CARBOHYD 561 561
 FT CARBOHYD 601 601
 FT VARSPLIC 218 242
 FT CONFLICT 19 19
 FT CONFLICT 55 55
 FT CONFLICT 71 71
 FT CONFLICT 89 91
 FT CONFLICT 94 94
 FT CONFLICT 540 540
 FT CONFLICT 668 668
 SQ SEQUENCE 794 AA; 91799 MW; 74DB3D236224499C CRC64;

Query Match 5.4%; Score 223; DB 1; Length 794;
 Best Local Similarity 20.1%; Pred. No. 0.035;
 Matches 162; Conservative 155; Mismatches 255; Indels 234; Gaps 40;

45 KEKKKNKDNKRDDE-----RNKTOEHLKEIKHIVKLEVGEEAVKKE-----AAEK 94
 76 KDSKNDKDVKREKEIRALLQERSTQKRIQDMESELEKTEAKLNAVREKTSISASNA 135
 95 LLEKVPSPDLEMYKAIGKIYVDGDIYKHISLEALSEDKKIKIDYKDALHEHYVA 154
 136 SLERLRELTFRANELLKK -FSEDG--HOKNMRALSLLEMLKLRN--KRETKMSMAYK 188
 155 KEKPEPLVIOSS- DYVENTEKALINVEYEGIKISRLDIKINO--DYQFLDVANTIK 211
 189 QEGME--LKLQATQKDLSEKGIYOLE--GKIVSIE-KEKIDKCEKTEKLEVIQDIS 242
 212 NASPSDQDGLFTNOLKEMPTDFSEF-----LEQN--SNEVOEYAKAFAYIIEFQH 262
 243 CASDQVECKVDIAQLEEDLKEKREILSLKQSLSEENTFSKQIEDLVKQ----- 293
 263 RDVQLYAPFAFNMYKDFNEQOINLSLE-ELKDQRMLSRYEKMEKIKOHYQWSDSLSEE 321
 294 -----QLETERDNLYSKDRERAFILSAMQILTERLALEREVEKLOQ----- 337
 322 GRGLLKQIDPIEKKDDIHSLSQEKELKRIQIDSSDF--LSTEKEFKLKQIDTI 378
 338 -----KELO-----SOSLLQOEKELSLRLOOQCLCSFOEEMTSKSNYKFEELKIAL 382
 379 RD-----SLSEEEK-----ELNRIQVDSNPLSEKKEFKLKIKIDQPY 419
 383 AELDAVOOKEQOSERLYKQLEEKRSSTAEOQLTRL--DNLLREKVELEKHAHAQAQI 438
 420 -----DINQRLQDTGGLIDSPSINDVKKQYKRIQIDIALHQSIGSTLVKIKIYLY 471
 439 LIAOEKYNDAQSLRDVYTAQLES-----VOEKINDTAQSL-----RVTYQLESQOEK 487
 472 EN--MININLTATLQADLVSDNTKINRGIFNEFKKNFKYSISNY-----MTVDIN-E 523
 488 NDTAQSLRDVYTAQLESQOEKYNDAQSLRDVYTAQLE-----SVQEKYNDTAQSLRDVSAQ 542
 524 RPAIDNERLKKRIQSLPTTRAGYIENGLILQIRNIGLEIKQVQIIOSEKEYIRIDAKV 583

Db 543 LESYKSTLKEIEDLK-----LEN--LTLOEKYAMKEKSEVEDYQ----- 580
 QY 584 PKSKIDITKIOEAQININQEWKALGLPYTKILIFNVNRYASINIVESAYILLNEMKNNI 643
 Db 581 -----QILTAE-STNE-----YARVQ-DLQNR-----STL 605
 QY 644 QSDLIKQVNTLVDSNGRFVFTDITLPINADQYHODEIV-EQVSKGLYVESRSL-- 700
 Db 606 KEKEIKETSSFLE-----KITDLKQOLRQODEDFRKLQLEKGRFAKENVMTTE 655
 QY 701 -----LHGSQGEVLKNDNS-----BEFIEFEGHAV-----DQVAGYLLD 734
 Db 656 LTMEINKRRLYEELYEYTKPPQOQDAFEAKQALLNEHGATQEQLNKIRDSYAQLLGH 715
 QY 735 KNSQDLVNSKKFIDIFKEGSNLTS 760
 Db 716 QNLKQKI-----KHVYKLDKNSQLKS 737

RESULT 8
 ID Y124.METUA STANDARD; PRT; 1075 AA.
 AC 057588;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-DEC-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN M0124.
 GN M0124.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 CC Methanococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA But C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kervilavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Uutterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.,
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RT Science 273:1058-1073(1996).
 RL
 CC -I- SIMILARITY: TO M.JANNASCHII M0124.
 CC -I- SIMILARITY: SOME, TO TYPE I RESTRICTION ENZYMES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U67469; AAB98104.1; -
 DR TIGR; M0124; -
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 1075 AA; 127796 MW; 4F765E1E0B52889 CRC64;

Query Match 5.3%; Score 218.5; DB 1; Length 1075;
 Best Local Similarity 21.1%; Pred. No. 0.074;
 Matches 192; Conservative 125; Mismatches 291; Indels 301; Gaps 49;
 QY 6 EFIRIVSMCSLVTAITTSAGPFIPLVQAGAGHDGVHVEKKEKKNKDNKRDERRKTKQ 65
 Db 277 EVISIKMSLSNMYIVAG-IYTTTIO-----KFOYRSKKKEKDNKKKDXD----- 322

```
OY 66 EHHLEIKMHVIEVKEGEEAVKKEAAE-----KLEKVPDVLDM 106
DB 323 DSDL-DLSKPIEELIKEDLKKEGKIKGLKDLILIAFYILKHEKNP-----DE 377
OY 107 YKAIGKITYVODITKHLSEALSEDDKKIDYKOLHEHYAAKEGEPVLVLOS 166
DB 378 YK-----KHI-----ENLKKLD-KD-----K 393
OY 167 SDDVNEPEKALNYYEIGKILSRDLISKINPYOKFLDVLTNTIKNASDDGDLFTNG 226
DB 394 KEETILN-----LONIKRHLILIDEHNRQYIGLGMKRIITPNAITFGTGT 443
OY 227 --LKEHPDTSVEFLEONSNEVOEFAKAPAYIEPHRDVLQLYPEAFNWMK--FNE 282
DB 444 PVFKNEKNFT--EFSYPEKGEFY-----LDVYFIGSINDKFTL--PLYQYVKEEDIS 495
OY 283 QINISLSEFLKQORMLSRYEKWEKIKOHYOH--SDLSSEGRGL-----KKIQPIE 334
DB 496 EGIQITLDE-----EDIEFIDEWIKRGEDINLFDKRLPKYINKSKITILN 542
OY 335 PKKDDIHSLOEKEKELLKRIQIDSSD-----LSTEE--KEFIKKIQIDRSLSEE 385
DB 543 PRIDRV-----AKIYDRIEDIEENEFKAMAVAVNRLGCVREFKALDKYIKKEFGDE 596
OY 386 EKELLNRIOVSSNPLSEKEK--EFLKKIKLIDIOPYD--INORLOD-----427
DB 597 AKKMAEVVVTYHNN--EKEKEIEYMKLKEKERNSDNEINQIIRREFLSSENKILL 653
OY 428 -----TGGLIDSPSIN--LDVKQYKRDIONID-----ALIHOSIG--STLY 465
DB 654 VTMMLLTG--FDAPRLKVMYLDKPLYGHRLLQAIARTNRPYDKERGLIVDSGLFKVLT 711
OY 466 NKIYLYENNINNLATLGADLVDSNDKINRGJFENRK-----KPKYSISSN 515
DB 712 EFMALYNMLAEELIRDEDFKNNLTISSIDE-----IFOEKLELWYKESLKNLK--INDE 763
OY 516 YMIVIDINERPAL-----DNERLKWRIOQLSPTRAGYLENGK-----LILQ 555
DB 764 DLSIVNTIKTLTKKNDFFNNNELKEL--DLIAFYADGKNARITKLIDLKAVIKLY 819
OY 556 RNIG-----LEIKDVQIKQSEKEYIRIDANVVPKSIDTKIQEA-----OL 597
DB 820 KALGYSYPOKIFYIEDIELLSFT--YAYLIKRLKPKKSNRKFWEELISFIHNKMLVDL 876
OY 598 NINQEMNKALGLPKYTKLITFNVHNRYASNIYESALILNEMKNNI--QSDLIKAVTVY 654
DB 877 TVIEEIN--LNPDDLDKILKENIGKREIKRAVANYFYL--KNSILDKOHDPITYK--927
OY 655 IVDGNGREFVTDITLPIIAE--QYTHODEIYOVHSKGLYVESRSLHGPSKGVLELN 712
DB 928 -----EILERLERLRKRWIMKRIDK-----IYNAIKNLMELN 962
OY 713 D-----SEGFIHERGHAVDYAGYLLDKNOSDL--VTNSKKFIDIFEKESNLTSYGR 763
DB 963 NYDKRIKGSISIERIKESISTYIGENILKQD-DIKLNLNTEFKLD--TKMQNLKLSK 1017
OY 764 TNEAEFFAE 772
DB 1018 LQRRKFKKE 1026
```

```
RESULT 9
V328_MYCGE STANDARD: PRT: 756 AA.
AC Q49419; Q49308; Q49320;
AT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MG328.
GN MG328.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
Mycoplasmataceae; Mycoplasma.
```

```
OX NCBI_TaxID=2097;
RN [1]
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bolt K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of mycoplasma genitalium";
RT science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 19-113 AND 155-272 FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bolt K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing";
RT J. Bacteriol. 175:7918-7930(1993).
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
CC EMBL: U39714; AAC71552.1; -
CC EMBL: U02203; AAD12492.1; -
CC EMBL: U02188; AAD12472.1; -
CC TIGR: MG328; -
CC KW Hypothetical protein; Complete proteome.
CC FT CONFLICT 52 52 T -> R (IN REF. 2).
CC FT CONFLICT 65 65 S -> W (IN REF. 2).
CC SO SEQUENCE 756 AA; 88407 MW; 1D3ADAEF0AB5C068 CRC64;
```

Query Match 5.3%; Score 218; DB 1; Length 756;
Best Local Similarity 21.6%; Pred. No. 0.051;
Matches 181; Conservative 145; Mismatches 268; Indels 244; Gaps 49;

```
OY 46 EKEKN-----KDEKRRKDE-----ERNKTOEHLKEIKHVIKVEGEAV 87
DB 61 EKEKSIYSEFFNMWIDLVKDKWKNVNLHYANQILKTRREEOQTUQTMAFOEVDQSYL 120
OY 88 KKEAAKLEKVP-----DYLEMKKATGKGIYVDSGDTKHSLEAL-----SPD 133
DB 121 TKEIKTGFQELKPSVYTAEDDKDE-----IKPEKTKOVSEELFNOPSEIINER 169
OY 134 KKKINDYKDALHBYVYAKGEPVLVLOSSEDE--VENNEKALNYYEIGKILSRD 191
DB 170 KKEVQVIFSTYDKV-----KEP-----EQFDFYSIENTKALN--PVHKITIQD 211
OY 192 ILSKINQPY--QKFLDVLTINKNASD--GODLLFTN-QLKEHPTDSVEFLEONSNE 245
DB 212 --QNDQDPVYVRIKLEQHPYTKKVDLDYNNKELLENNADLK--QIDDLKENMND 264
OY 246 VOEVFAKAFAYIIEPHRDVLQLYAPEAFNWMKKEOEINLSLEELKQORMLSRYEKME 305
DB 265 --QIF-----DL-----EQET-----DCLK--RLLSF--E 283
OY 306 KIKQ--HYQMSDSLSEGRGLLKKLO--IPIEPKKDDIHSLS--QEEKELLKRIQIDS 359
DB 284 KSKHLHTTKLQDLDLQENNDIYEQLOKMKRYALPISDEVNEBELNKQKALLSD-QDQA 342
OY 360 SDFLSTEEKEFLKQIIDIRDSLSEEEKELNRIQVSSNPLSEKEKFLK--KKLDI 416
DB 343 LKKSSNVOOQIALLPV-LNNOINELQNDLITAREANORLIDVEQBNDFLKNELKHLHON 401
OY 417 QPDIYQRIQDQNG-----LIDSPSINLDVRKQYKRRIOINIDLHSGISSTLYLNKLYLE 472
```



```
Db 402 TSNDENEYDLDLNOYELLPEBNETKPD-KIOVQOALNLD---YOKITSALKHE----- 452
OY 473 MNMINNLTATLADLVSTDNFKINRGIEFFKKNFKYSSIMYIVDINERPALD----- 528
Db 453 ----NDV-----LIDIEEWTSRKDNDFNMTKNSFE-----EQKALDEKLN 489
OY 529 -----NEMLKRI-QASP-DTRAGYLENGKILLQNRIGLEIKDVOILIOSEKEIYRIDAK 581
Db 490 GLTIQNOOOLQKIALELEBMEENKSNLNTNOLV-----NLOOQOLKDSQNLFN----- 535
OY 582 VVPSKIDTKIOEAOALINQEMN-----KALGLPKYTKLITFVNRHNRASNIYESAYLTIN 637
Db 536 -VADOKLAT-LEEVNLAINEKINLEDELSG-----SENSNNLAK--LQADHETLO 583
OY 638 EMKNNIOSDLIKKYTNVLVDNGRFPVTDITLPNIAEQYTHODEIYEQVHSGKGLVPSR 697
Db 584 ESYGKLTKDFEKLKKNKLNDA-----EQYDOLLSAFETNSE---LEKAK 626
OY 698 SILHGPSKGYELR-----NDSEGFIERGHAVDDYAGYLLDKNOSDLVYNSKKFDIFKEE 754
Db 627 QSLASDSENNOLKQOINSLNARKEL-----OETPVTSDEHDELET-----LKIEKEQ 676
OY 755 GSNLTSGRTNEAEFFAFRLMSTHDAERLKVKOKNAPKTFEOFINDQIK---FLINS 809
Db 677 -LFIENQALQLOQYFNDI-----SANTOEIKASDEKDEVEIKPKIKKRDVYQIN 728

RESULT 10
ID Y109_YEAST STANDARD; PRT: 1679 AA.
AC P40457;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE HYPOTHEICAL 195.1 KDA PROTEIN IN DNA43-UB11 INTERGENIC REGION.
GN Y1149C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RA Bartell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Goules S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z38059; CAAB6129.1; -.
DR PIR: S48385; S48385.
DR SGD: S0001411; MLP2.
KW Hypothetical protein.
SQ
SEQUENCE 1679 AA; 195141 MW; 298950CC52202D8F CRC64;
```

```
Db 644 IYKSOCKKKTTLDEFFENFKLAKEMLEEAIDL-KALEKOKSVSPYTHVEKERA 702
OY 93 EKLLEKVPDYLEWYKAIGRTYIVDGDITKHSLEALSSEDKKIKD-----IYKDAL 147
Db 703 STELSQ-----SRIKTISLEYEISIKLKETAFTASFIPTKEST 738
OY 148 HEHYVAKESYEPVLIQSSPDYVENTEKALNVYEIGKIISR-----DILSKINOPQK 202
Db 739 RDEQCCKEKKELOMRLEKESE--ISHNENKMDFFSSKEGOYKAKIKELNNLERLSQLOS 796
OY 203 FLVDLNTTKNKSADS-----DGODLFTNQKE-----HP 231
Db 797 KIOEESTIRSCDSQOLKRAQNTIDDETKMKMSLTELSENKETTETKLSSEINDEKLEIK 856
OY 232 TDESVEFLEQNSNEVOEYFAKAFAYIEPQHR---DVLQYAPAFNYMKENE--QGIN 286
Db 857 TKFOYKFLDQNSD-----ASTLEPTLRKELEQIOVOLKANQSOIQAYEETISSNE 906
OY 287 LSLFEIKQRM-----LSRYEKWEKIKQHYQMSDSLSEEGRLKKIQIPIEP 335
Db 907 NALIEKLAKTEKENYDAKTELEKKEKWAR-----EEDLSRL-RGELGEIR-ALQP 956
OY 336 K-KDDIHSLSQBER--ELKRIQ--IDSSDFLSTEEKFLKKIQIDIRDSLSEEEK--- 387
Db 957 KIKGALHFVQOSKRLRNEVERIQMKIEIKMST-IYQLCKKEMSGYSTMKENKDL 1015
OY 388 ELNRIOVDSSNPULSEKEKEFLKKLKDIDPYDINQRLQDTGGLIDPSIN-LDVNRQYK 446
Db 1016 ELVIRLEKMDAQAE-----LTKTK-----SLYSADQLDKHERKWEKKADE 1061
OY 447 RD-IQNTDALLHQSTGLYKTKIYENMNINNLATIGALVYSTDTKTRNGIFNEFK 505
Db 1062 RELISNIEQT-----ESLRVEN-----SVLIRKVDYFANNQDKHLK 1099
OY 506 KNFYSSIMYIVDINERPALDNERLKWRIQSLSPDFRAGYLENGKILLQNRIGLE--IK 563
Db 1100 ---LVSLFSNLR-----HERNSLETKLTCKRELA-----FYKQNDSELEKTIN 1140
OY 564 DVQIILQ-SEKEYIRIDAKVVPKSKIDTKIOEAOALINQENKALGLPKYTKLITFNVHN 622
Db 1141 DLQRTQTLSEKEY-QCSAVIIDEFKDITK-EVTQVNIILKE--NNAI-LQKSLKNTYEKNRE 1196
OY 623 RYA-----SNIV-----ESAYLLNEMKNNTQSD 646
Db 1197 IYKOLNDROEISRLQRLDITKQOVSTNSKNKILVESEMQQKQYODLSQQKDAQKK 1256
OY 647 LIKRYTNVLVDNGRFPVTDITLPNIAEQYTHODEIYEQVHSHK----- 689
Db 1257 DIEKLTNFSIDLKGLKLSAEANANDLENKF--NRKKQAHEKLDASKKQOALTNELNE 1313
OY 690 -----GLVPEESRILLHGPSKGVLELNDSEGFIEFGHAVDDYAGYLLDKNOSD 739
Db 1314 LKAIKDKLEODLHEENAKVUIDLDTFKLAKHELOSSEVSRDHE-----KDY-----FTLME 1363
OY 740 LVYNSKKRPIDFKEGSLTGYGRTNEAEFFAFRLMSTHDAERLKVKONAPKTFQFI 799
Db 1364 EIESLKELOLQFKYANS-----SSDAFEKLVNNEKBERD-RII 1400
OY 800 NDOIK 804
Db 1401 DERTK 1405

RESULT 11
ID MYSJ_DICDI STANDARD; PRT: 2245 AA.
AC P54697;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN II HEAVY CHAIN.
GN MYOJ.
OS Dictyostelium discoideum (Slime mold).
```


OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AX3;
RX MEDLINE=96215148; PubMed=8636147;
RA Hammer J.A. III, Jung G.;
RT "The sequence of the dictyostelium myo J heavy chain gene predicts a
RT novel, dimeric, unconventional myosin with a heavy chain molecular
RT mass of 258 kDa."
RL J. Biol. Chem. 271:7120-7127(1996).
RN [2]
RP SEQUENCE OF 1-1021 FROM N.A.
RX MEDLINE=97039016; PubMed=8884597;
RA Peterson M.D., Urzoste A.S., Titus M.A.;
RT "Dictyostelium discoideum myoj: a member of a broadly defined myosin
RT V class or a class XI unconventional myosin";
RL J. Muscle Res. Cell Motil. 17:411-424(1996).
RN [3]
RP SEQUENCE OF 182-298 FROM N.A.
RX MEDLINE=95023928; PubMed=7937787;
RA Titus M.A., Kuspa A., Loomis W.F.;
RT "Discovery of myosin genes by physical mapping in Dictyostelium.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9446-9450(1994).
CC -I- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN AND HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -I- SUBUNIT: HOMODIMER.
CC -I- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -I- SIMILARITY: CONTAINS 3 IQ DOMAINS.
CC -I- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage, by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U42409; AAA85186.1; -;
DR EMBL: J35322; AAA79858.1; -;
DR HSSP: P08799; 1MND.
DR DICTYDB: DP01095; myoj.
DR InterPro: IPR002710; DIL.
DR InterPro: IPR000648; IQ.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF01843; DIL; 1.
DR Pfam: PF00612; IQ; 3.
DR Pfam: PF00063; myosin_head; 2.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 1.
DR ProDom: PD000376; DIL; 1.
DR SMART: SM00015; IQ; 3.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 3.
KM Myosin: ATP-binding; Actin-binding; Calmodulin-binding; Repeat:
KM Coiled coil.
FT DOMAIN 1 809 MYOSIN HEAD-LIKE.
FT DOMAIN 824 851 IQ 1.
FT DOMAIN 872 901 IQ 2.
FT DOMAIN 943 972 IQ 3.
FT DOMAIN 973 1812 COILED COIL (POTENTIAL).
FT DOMAIN 1813 2245 TAIL.
FT NP_BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 669 749 ACTIN-BINDING.
FT DOMAIN 191 191 L -> F (IN REF. 2).
FT CONFLICT 284 284 A -> T (IN REF. 2).
FT CONFLICT 291 291 G -> R (IN REF. 2).
FT CONFLICT 332 347 NKGGEIEGVSDDEH -> IEMFELKIVRMKS (IN
FT REF. 2).
FT CONFLICT 550 550 N -> K (IN REF. 2).
FT CONFLICT 865 866 HH -> QQ (IN REF. 2).
SQ SEQUENCE 2245 AA: 258478 MW: 615E5EFID1AB45BR CRC64;

Query Match 5 2%; Score 215; DB 1; Length 2245;
Best Local Similarity 19.4%; Pred. No. 0.24;
Matches 169; Conservative 147; Mismatches 274; Indels 280; Gaps 40;
QY 50 NKDENKRRKDEE-----NKTOEE--HKEIKKHIVK--IEVKGEE-- 85
DB 1123 NKEPRLRSEGRDTNNTNNOLEIQOLKANSTLEEDVFSLSGIRDNREQLRDEENQL 1182
QY 86 -----AVKKAEEKLE-----KVPDYLEMYKAIIGKKIYVDS 119
DB 1183 IKERLDSLGQSSQFQSGALERQOLEIOLVQEOSQOLIKLSEKLGSEERAKQIOMLEL 1242
QY 120 DITKH-----ISLEALSEDKKIKIDYGRDALLHEHVYAKEGEPELVQSSDVEYNT 174
DB 1243 ELTDHRSKLOIQOLTEQSEKIKKLKG--LEETODEKKOLOELERIKOSKOSVEDE 1299
QY 175 EKALNYYEIGKILSRDILSKINOPYOKFLDVINTIKNASDSG----- 218
DB 1300 KNSLITQLTVKFEPSQVSTNVSHQKIKITLSTIELKSKIGKIQAEQKNKDEIRKI 1359
QY 219 -----QDLFTNQLKEH-----PTDFSVFELPQNSNEVOYAKAFAYIEP 260
DB 1360 QFELNDQKQFTRQTKFESDLSQSQSIDRPKSEITIHSLERTETLKSDFERV--QOSLQ 1418
QY 261 QHRDVQLVAPFAFNVDKN--EQEINLSLEELKD----- 294
DB 1419 QERDCQ-----YKDTINRLENEVK-QLTQKEFEFEFFVAKQNSNQTESYLLK 1469
QY 295 -----QRLSRYEK-WEKIKOHWSDSLSEGRRLKLQIPLEPKDDIHSLSQ 347
DB 1470 EYTTQMQNQSRLEKELEKKQHTTRIDF-RDELKQQLQLOQHQHQSSTQL--LAQN 1526
QY 348 EKELLKRIQI--DSSDLSTEKEF--LKLQIDIRDSLS-----EEKELLNRIQVD 396
DB 1527 ELELRKKEIKYKERGHETSQKQDQFMHEIQSLRTNNQDLSQDYEQEKKL-KDKLS 1585
QY 397 SSNPLSEKEKEFLKKLKDQPY-----DINQQLQDTGGLIDPSNLVDVRKQ 444
DB 1586 SSKQEAQQQRESEIILKMAELSAIKQHSQWENSTDMKQKQNE--LIESAL--YKQ 1639
QY 445 YKRDIQNTDAL-----LHQ-----SISSTLYNK--I 468
DB 1640 LLOQTSTIDSTIKENKESIKLQQLQETSNOQLHQLEELNSMKQNSQLSTEDSKQLNQ 1699
QY 469 YLYENNMNNILATLGADLVDS--DNTKIRNGIFNEFKKKNKYSISSNY----- 516
DB 1700 LIQENQQLKSVTNEISKQLDDAVENQKINNITKEDEIKSKRMSVELQOHIDEKQOEIQ 1759
QY 517 ----MYDINERPALDNERLKWRQ-----LSPTFRAGY--LENGKLLQORNG 559
DB 1760 QLOSTIAQLKQOQOQSETRLEKEIQQKREREPQMKLVESSTRKLVHMLED-RMELVKNV- 1817
QY 560 LEIKDVQIIKQSEKEYI-----RIDAKVYPKSIDTKIQEAOVLNIQEW----- 603
DB 1818 MEIID---YKETEWEKRLARLAGKEIDTKLISDLSCKLEHSLG--SQMFFHIDVWCP 1873
QY 604 ----NKAIGLPKYTKLITENVN---RYASNIVESAYLILNEKNNIQSDLIKVKYTV 654
DB 1874 YERDSKGIYGIIRSIYDFTIKNFDDVLLSYLLACCSITLFLYKKNLVKHL----- 1926
QY 655 LVDSNGRQVFTDI-----TLPNIAEQYTHQ 679
DB 1927 ----NGANSIMPIPTLGLDEELNERLSHQ 1952
RESULT 12
YDBE_SCHPO
ID YDBE_SCHPO STANDARD; PRT; 1957 AA.
AC Q10411;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, last annotation update)
 DE HYPOHETICAL 222.8 KDA PROTEIN C1F3.06C IN CHROMOSOME 1.
 GN SPAC1F3.06C.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomycetes.
 OK NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; 270690; CAA94624.1; -
 KW Hypothetical protein.
 SO SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

Query Match 5.1%; Score 212.5; DB 1; Length 1957;
 Best Local Similarity 18.8%; Pred. No. 0.25;
 Matches 188; Conservative 155; Mismatches 342; Indels 317; Gaps 44;

QY 59 FERNTOEHLKEIMKHIVKIEVKEEAVK-----EAAEKLEKYPDYLEMYKAIG-GK 113
 DB 650 DKESTFLSDENNDDTKLTKLEESNKSILTKQEDVDSELEKNQOTLKEDLRKSEALRFSK 709
 QY 114 I-----YIVDDITKHISLEALSED-KKIKIDYGDALHEHYVYAKEGEPEV----- 161
 DB 710 LEAKLREVIODLKKGHELEQNRDLHSSLSDAKNTAAILSSSELTSESDYKRLTANVE 769
 QY 162 -----LVYOSSEDEVVENTEKALNYYE----- 183
 DB 770 TLTQDSKAMKOSFTSLVNSYOSISMLYHELRDHYNMOSQNTLTLESSEKLTDCENLTQ 829
 QY 184 -----IGKILSRDI-----LSKINOPYQKFLDYLVNTIKNASDGDODLLFTNQL 227
 DB 830 QNMTLIDNVQKLMHRYKVOESKVSLEKSLSLDKNLKRSLSVAALSDNDQIL--TQL 887
 QY 228 KEHPTDFSEVEPLEQNSNEVEQVFAFAFYIEPQHRDYLQALPAFNYMKF----- 280
 DB 888 AELSKNY--DSLQESADLNSGLKLEA-----EKQLHTNEHELHRLDKLTGKLTIE 939
 QY 281 -----NEOET-NLSLEELKQORMLRYE-----KWEKIKOHYQHWSDSL 318
 DB 940 ESKSSDLCKKLTARQEEISNLKEENMSOQAITSVKSLDETLSKSKLEADIEHLKNVY 999
 QY 319 SE--EGRGL-----KLIQPIEPKK--DDIILHSL----- 344
 DB 1000 SEVEYERNAALASNERLMDLKNNGENASLQTELEKRAENDDIQSKLSVSSYEENLL 1059
 QY 345 --SOEEREL-----LKRIO-----IDSDFLSTEEKEFLKLT----- 374
 DB 1060 LISSQTNKSLDEKTNQKLYIEKNVQKLLDEKQNRNVELEELSKGKLGEEANAQIKDELL 1119
 QY 375 -----QIDIRSLSEEEKELNLRIQVDSN--PLSEKEKEFLK 410
 DB 1120 ALRKSKKHQDLCAFYVDLKEKSDALEQTLTEKNEMLIYLSQSSNNNEALVEERSDLAN 1179
 QY 411 KTKLDIQYDINQRLQDGGGL-----DSPSIN--LDYRK-----QYRKDIQNIAL 455
 DB 1180 RLS-----DKKSLSDSDNNIVSIRSLVARNDELDTLKKDKDLSLSTQYSEVCCDRDDL 1233
 QY 456 LHSQIG-STLVNK--IYLYENNNINNLATGADLVDSQDNTKIRGIFNEKK----- 506

DB 1234 LDSLKGESEFNKRYAVSLRELCTKSEIDVPV--SEILD--DNFVNAAGNFSLSRLTVLSL 1290
 QY 507 -----NEKYSISSNYM-----IVDINERPALDNERLKWRIQSLPPTRAGY 546
 DB 1291 EKVLDAPFNQVNRKKNKHELOLRLLTTDAEFTKRYVADLEK--LOHEHDDMLIQ-----RD 1341
 QY 547 LENGKLIILORNIIGLEIKDVOIITKQSEKEYIRIDAKV-----VPSKIDTKIQEQNLIN 600
 DB 1342 LE-----KALKDESEKNFLRKEAEWENTENIHSLSEKKEETKEIAELSSR 1384
 QY 601 QEMNKALGLPKYTKLITTFVNHRYASNTV---ESAVILINENKN--IQSLDI--KVT 652
 DB 1385 LEDNQATNKKLNQDLHQLQETRLKEDVLKESKESLITSLSELSNOROKESSLLDAKNE 1444
 QY 653 NYLVQNGRFRVFTDITLPNIAEQYTH--ODEIYE-----QVHSKGLVPEERS 698
 DB 1445 EHMUDTSR---KNSLMEKIESINSLLDKKFFELASANEKGLAOKLHSELSLMEYTK 1501
 QY 699 ILLHGPSKGVLELRNDSGEFHEFGHAV---DDVAGYLDKMQ--SDLVTSKKRFTIDFK 752
 DB 1502 SOLQEAKEKIQV---DESTIOELDHEITASKNNYEGKLEMDKDSIIRDLSENIQNLNLA 1558
 QY 753 EEGSNLTSYGRTEAEFFAEAFRLHSTDAERLKVOKNAPK 794
 DB 1559 EKSNAVKRLSTKESEELIQFNSRLADLEYHKSQVSELSGRSK 1600

RESULT 13
 MPL1_YEAST
 ID MPL1_YEAST STANDARD: PRT: 1875 AA.
 AC 002455:
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MYOSIN-LIKE PROTEIN MLPI.
 GN MLPI OR YKR095W OR YKR415.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 CC NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=S288C.
 RC MEDLINE=93247549; PubMed=8483450;
 RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
 RT "A new yeast gene with a myosin-like heptad repeat structure.";
 RT Mo. Gen. Genet. 237:359-369(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=94205265; PubMed=8154186;
 RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
 RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
 RT "The complete sequence of a 15,820 bp segment of Saccharomyces
 RT cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
 RT new open reading frames.";
 RL Yeast 9:1349-1354(1993).
 CC
 CC -I- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
 CC REPAIR.
 CC
 CC -I- SIMILARITY: SOME, TO THE TPR ONCOGENE.
 CC
 CC -I- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L01992; AAA34783.1; -
 DR EMBL; X73541; CAA51948.1; -
 DR EMBL; 228320; CAA82174.1; -
 DR PIR; S38173; S38173.

DR SCD: S0001803; MLI.
 KM Colled coll; DNA repair.
 FT DOMAIN 69 487 COILED COIL (POTENTIAL).
 FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
 FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
 FT CONFLICT 301 301 R -> A (IN REF. 1).
 SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C9066867 CRC64;

Query Match 5.1%; Score 211.5; DB 1; Length 1875;
 Best Local Similarity 19.9%; Pred. No. 0.26;
 Matches 193; Conservative 178; Mismatches 330; Indels 269; Gaps 51;

OY 2 NIKKEFIYI-----SMSCYVTAI-TLSCGVFPIVQAGAGHDDVGHVKEKKNNDENR 56
 DB 765 NLKQLELALSPKSDRLRWITOLTLQ-----KEREDLLEETTK 803
 OY 57 KDEBNKTOEELKEIMK-----HIVKEVKE-----EAVKEAEKLEKY 99
 DB 804 SQCKIDELEDALSELKETSOKDHIKOLEEDNNSNIEMYNKIEALKKD-YESVITSV 862
 OY 100 PSDVLEMTKALGKIYIVDGI-----TKHISLEALSEDKKIKIDYGRDALLHENY 151
 DB 863 DSKQPDIEK-LOKYKVSLEKEIEEDKIRLHYVMEDETINDSLR-KELEKSKINLTDAY 920
 OY 152 VYAKGEYEVLIQSEDEVVENTEKALNYYEIGKILSDILSKINOPQKFLDVNTIK 211
 DB 921 SOIKK-----YKDLVETTSOSL-----QOTNSKID--ESERDFTNQIK 956
 OY 212 NASD-----SDGODLL-----FTNQL-----KEHPDFS--VEFLRONSNEVOEVA 251
 DB 957 NLDEKTSLEDKISLKEQMFNLNDELQKKMEKEMKADFPKRISILQNNKKEVAVAS 1016
 OY 252 KAFAYITPQH-RDVQLYAPPA-FNYMDKFNED-EINLSLELKDQ-----RML 298
 DB 1017 EYESLRSKIQNDLDOQTLYANTAONNVEQLKHAADVSTIELRQHLTYGQVKTLM 1076
 OY 299 SVEKMEKIKOHYHMS-----DLSSEGRGLKLIKIDPIEPKKD--- 338
 DB 1077 SMDQLEMLKENEKMSKESLEFOLDLSNRIEDLSQKNKLIDVOIITYAADKEVNN 1136
 OY 339 -----DIHLSQOEKELLKRIQIDSSFLSTEEKFLKLOI-DIRDSL--SEER 387
 DB 1137 STNGPGLNNILITLRRERDILDTKYVAERDAKMLRQKISLMDVELQDARTKLDSRVER 1196
 OY 388 E-----LLNRQVDSNPL--SEKEKF--LKKLIDIP--- 418
 DB 1197 EHHSSIIQOHDIMEKLNQNLRESNITLRNLENNNNKKELQSELDLQONVAPIES 1256
 OY 419 -----YDINORLODTGGLIDPSJINLDVQRQ---YKRDIONIDALLHQSIGSTLYNKI 468
 DB 1257 ELTALKYSQKEQOE-----LKLAKEEVHMKRRSODI-LEKHEQSSSYDETK 1304
 OY 469 VLYENMNINILATL-----GADVSTDNTRKINGIFNEFKKNFKYS-ISSNYIVDI 521
 DB 1305 -----ESEIENLKEELENERQGAFAE-----KPNR-LRQOERLRTYSKLSQDSDILEY 1354
 OY 522 NE-----RALDNERLKWRIQLS-PTPRAGYLENGKILIORNIGLE-IKDVQITIKOSEKEY 575
 DB 1355 NSLRDAKVNLEN-----SLSEANARIEELONAK-VAQGNNOLEAIRKIQ--EDAEKAS 1404
 OY 576 IRIDAKVVPKSIDTKIOEALN-----INQENMKALGPKYTKLITPVNHNKYAN 627
 DB 1405 RELQAKL-----EESTTSTESTINGLNEETITTYKELEKORQIQOOLQATSAEDQND-LSN 1459
 OY 628 IVESAYLILNEMKNNIQSLIKKVTNYLDVNGRFV-----TDITLPINAEQY--T 677
 DB 1460 IVES-----MKKSFEEDKIKIFIKETQEVNEKILEAQERLNOPSNIMEELIKKKWESE 1512
 OY 678 HODEIYEQVH-----SKGLVPESSSILLHGPSKGVLELRANDSEGTIHEFGNAVDYAGY 731
 DB 1513 HDEVSOKIREAEALKKRIRLPTEEKI-----NKIIEKK--KEELEKEFEKVERERIKS 1565

OY 732 LLDKNOSDLVTNSKRFIDIFKEGNSULTSYGRTEAEFPALRMLHSTDAERLKVOKN 791
 DB 1566 MEQSGEIDVYLKOLEAKVOEKOLENEXNKKLQLELDQVHSSHISIDDERDKLAETI 1625
 OY 792 AKPTQFQID 801
 DB 1626 SRLREEFNNE 1635

RESULT 14
 RBP2_PLAVB
 ID RBP2_PLAVB STANDARD; PRT; 1251 AA.
 AC 000799;
 DT 01-APR-1993 (Rel. 25 Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE RETICULOCYTE BINDING PROTEIN 2 (FRAGMENT).
 GN RBP2.
 OS Plasmodium vivax (strain Belen).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=31273;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9215338; PubMed=1617731;
 RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.:
 RT "A reticulocyte-binding protein complex of Plasmodium vivax
 mezoites."
 RL Cell 69:1213-1226(1992).
 CC -I- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
 CC HUMAN RETICULOCYTE CELLS.
 CC -I- SUBCELLULAR LOCATION: MEMBRANE-BOUND (PROBABLE).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sb.ch/announce/>
 CC or send an email to license@isb-sb.ch).
 CC
 CC EMBL: M88098; AAA29744.1; -
 KW Malaria; Receptor; Membrane.
 FT NON_TER 1
 FT NON_TER 1251
 SQ SEQUENCE 1251 AA; 143741 MW; 54BA51C7404AC572 CRC64;

Query Match 5.1%; Score 211; DB 1; Length 1251;
 Best Local Similarity 20.7%; Pred. No. 0.17;
 Matches 171; Conservative 136; Mismatches 286; Indels 232; Gaps 44;

OY 45 KEKKNKDEKRRKDEBNKTOEELKEIMKHIVKEVGEAVKKAELKEKVPDVL 104
 DB 6 KYVDTSFDEKKSIEKAYEKMGTNLKELEK-----MDKEKNEKEVEEAOI----- 52
 OY 105 EHWKALGKIYIVDQITIHISLEALSEDEKKKIKIDYGDALLHENYVAKGEYEVLYI 164
 DB 53 --YKRI-----FLDHVY-----NLNDEVEKSKIVMEKTEL-----YKKEDE--IK 90
 OY 165 QSEDEVVENTERKALNYYEIGKILSRDILSKINOPQKFLDVNTIKNASDSDGODLLFT 224
 DB 91 QKTNEKQODTS--NYYTYEQQVYSAQSOKAKIEQ-----FINATITTKGTSDT-SQIINEL 143
 OY 225 NOLKEHPTDFSYEELEQNSNEVOEFAKAFAYIIEPQHNDVQLVAPAFNYMDKFNEDQ 284
 DB 144 ESIIKEE-VHKNLQOLVQOESNMEEMKQILSM-----KDLILNNSETIAKISNNTON 196
 OY 285 INLSLELQDQRLSLRYEKMEKIKOHYHMSDSLSEGRGLKLIKIDPIEPKK-DDLIHS 343
 DB 197 ALGFRENAK-----TKLNTDDELQRYA-----AMIEAKAHKNNIDIALDAQIDTEVSK 247
 OY 344 LSGEKKELKRIQIDSSDFLSTEEKFLKLIKIDIRDSLSEEEKELNRIQVDSNPLSE 403

```

Db 248 IEQINREIMNK-----KDEIKSYLSEI-----KEYKDCCTEISNSKRG 286
Qy 404 KER-EPLKALKLIDOPDINORDTGLDPSINLDVR--KOYKRDIONDALHLOSI 460
Db 287 KDKIEPLEKFK-----PNE-----ESNSKNVINEINENIRSEQYLKIDIE--DAEKQAST 335
Qy 461 GSTLYNNKTYLENNNNINL--TATGADLVSDNDNRKING--IFNEFK-----505
Db 336 KVELFHK--HEPTTISIFKESEILG--VETKSOKKINAEDIMEIEEHNSEIOTOV 388
Qy 506 KNFYSISSNMYIADINERPALDNERLKWRIQLSPDTRAGYLENGKLIORN--IGLEI 562
Db 389 KGFENLNK-----LNEPHYDNA-----EDELNDKST--NAKYLIEINLESVKHNL 434
Qy 563 KDOVITKO-SEKEYIR--IDAKVVPK-----KIDTKIOQAOLINQEWKALGLPKYT 613
Db 435 SEINIKOGGEIKYSKAKDIMOKIKATSENTAEKLEKVKDQSN-----479
Qy 614 KLIFVNNRNASNIVESAVILNEMKN-----NIOSDLIKVTNYLVD-----657
Db 480 -----YVNLNOITTERNLITE-KNRLNGIDSTITNIGALKESKNGNEIGLEKLE 531
Qy 658 --GNGREFVETDITLPNI-----AEQYTHODEIVEOVHSGLYVESRSIL 700
Db 532 EIGNRRLKVDITKKSINSTVGNFSLEFNNEDLNQYDENKNKINDYENKMG-----581
Qy 701 LHGSKVELANDSEGTHERFGHAVVDYAGYLLDKNOSDLVTSNKKFIDI-FKEGSGNLT 759
Db 582 -----EYNEPEGSINKISENLRN-----ASENTSDY--NSAKTLRLAEKREKVL- 625
Qy 760 SYGRNEAEFEAPRLMHSSTDAERLKYOKNAKPTQFINDOIK 804
Db 626 ----LNKEE--EANKYLIDVKKVESFRFIFNMKESLDKINEMIK 663

RESULT 15
MY6_MESAU STANDARD; PRT; 1939 AA.
ID MY6_MESAU STANDARD; PRT; 1939 AA.
AC P13539; 060562; 13, Created
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE ALPHA ISOFORM (MYHC-ALPHA).
GN MY6.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F1B; TISSUE=Liver;
RC MEDLINE=95115033; PubMed=7815459;
RA Wang R., Sole M.J., Cukerman E., Liew C.-C.;
RA "Characterization and nucleotide sequence of the cardiac alpha-myosin
RT heavy chain gene from Syrian hamster."
RT J. Mol. Cell. Cardiol. 26:1155-1165(1994).
RN [2]
RP SEQUENCE OF 1630-1939 FROM N.A.
RP MEDLINE=86205859; PubMed=3458174;
RA Liew C.-C., Jandreski M.A.;
RA "Construction and characterization of the alpha form of a cardiac
RT myosin heavy chain cDNA clone and its developmental expression in the
RT Syrian hamster."
RL Proc. Natl. Acad. Sci. U.S.A. 83:3175-3179(1986).
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.

```

```

CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY (BY
CC SIMILARITY).
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -1- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L15351; AAB59701.1; -.
DR EMBL; M12995; AAA37081.1; -.
DR InterPro; IPR000048; IO.
DR InterPro; IPR002928; Myosin_head.
DR InterPro; IPR001609; myosin_tail.
DR Pfam; PF00612; IO; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IO; 1.
DR SMART; SM00242; MSc; 1.
DR PROSITE; PS50096; IO; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family;
KW Calmodulin-binding.
FT DOMAIN 1 782 MYOSIN HEAD-LIKE.
FT DOMAIN 783 812 IO.
FT DOMAIN 842 1939 COILED COIL (POTENTIAL).
FT DOMAIN 1439 1443 POLY-ALA.
FT NP_BIND 178 185 ATP (POTENTIAL).
FT DOMAIN 657 679 ACTIN-BINDING.
FT DOMAIN 759 773 ACTIN-BINDING.
FT MOD_RES 129 129 METHYLATION (TRT-) (POTENTIAL).
FT MOD_RES 697 697 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 707 707 ALKYLATION (SH-2) (POTENTIAL).
FT CONFLICT 1633 1633 O -> L (IN REF. 2).
FT CONFLICT 1651 1651 H -> Q (IN REF. 2).
FT CONFLICT 1686 1687 H -> DV (IN REF. 2).
FT CONFLICT 1693 1693 V -> G (IN REF. 2).
FT CONFLICT 1844 1844 K -> R (IN REF. 2).
FT CONFLICT 1879 1879 A -> T (IN REF. 2).
FT CONFLICT 1885 1885 E -> Q (IN REF. 2).
FT CONFLICT 1907 1907 E -> V (IN REF. 2).
FT CONFLICT 1928 1928 D -> N (IN REF. 2).
FT CONFLICT 1933 1935 OKM -> KR (IN REF. 2).
SQ SEQUENCE 1939 AA; 223626 MW; DBC8297DFE83115A CRC64;

Query Match 5.1k; Score 210; DB 1; Length 1939;
Best Local Similarity 18.6k; Pred. No. 0.31;
Matches 159; Conservative 159; Mismatches 261; Indels 256; Gaps 36;

Qy 29 PLVYGAGHGVDGNHVEKEKNKDNKKTQEBHLEIM--KHIVKEVGGEA 86
Db 840 PLKSAFTEKEMA-NMKEDEFRVAKESLEKSAARRKLEKKVSLQEKNDLPQVQAEQD 898
Qy 87 VKKAAEKLEKVDVLEMTKAIGKIYVDGITHISLEALSEDKKTKIDYGDAL 146
Db 899 NLNDAEER-----CDQLIKKKIQLEA-----KVE----- 923
Qy 147 LHEHYVAKGEYEVLVIOSESDYVENTEKALNYVEIGKILSDILSKINPQKFLDV 206

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 2, 2001, 13:49:05 : Search time 1166.37 seconds
(without alignments)
22387.614 Million cell updates/sec

Title: US-09-747-521-1
2430
Perfect score: 1 atgatatataaaaaagaatt.....agttcattatataactataa 2430
Sequence:

Scoring table: IDENTITY_NUC
Gap 10.0, Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estfun:*
2: em_estlin:*
3: em_estl:*
4: em_estom:*
5: em_estipl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_estl:*
11: gb_estl2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	103.2	4.2	1337	5	BF630719
C 2	102.6	4.2	1025	13	CNS014J2
C 3	93.8	3.9	1101	13	CNS0021J
C 4	92	3.8	1059	13	CNS0022B
C 5	91.8	3.8	1036	13	CNS031MT
C 6	91.6	3.8	1036	13	CNS00599
C 7	90.8	3.7	1885	10	BE420745
C 8	89.8	3.7	1101	13	CNS005EV
C 9	89.2	3.7	1135	13	CNS033GO
C 10	88	3.6	1001	13	CNS0155H
C 11	87.4	3.6	576	13	CNS035N7
C 12	87.4	3.6	1101	13	CNS0039C

C 13	86.8	3.6	838	5	BF629227	BF629227 HVSMB001
C 14	86.8	3.6	963	13	CNS0044L	AL054918 Drosophila
C 15	86.6	3.6	1020	13	CNS038G9	AL232578 Tetradon
C 16	86.6	3.6	1151	11	BG309087	BG309087 HVSMB000
C 17	86.4	3.6	908	13	A2548467	A2548467 ENTER30TR
C 18	86.4	3.6	1101	13	CNS00K85	AL077453 Drosophila
C 19	86	3.5	520	13	CNS02LKJ	AL202924 Tetradon
C 20	86	3.5	1223	13	B12981	B12981 T24D11-Sp6
C 21	85.8	3.5	879	13	CNS01RFG	AL147405 Anopheles
C 22	85.8	3.5	982	13	AQ325799	AQ325799 nbx00021B
C 23	85.6	3.5	1101	13	CNS000SX	AL050813 Drosophila
C 24	85.4	3.5	1101	13	CNS0153V	AL104965 Drosophila
C 25	84.8	3.5	471	10	AI404883	AI404883 GH24733.5
C 26	84.8	3.5	1101	13	CNS016OP	AL107011 Drosophila
C 27	84.6	3.5	834	13	B12387	B12387 F21E20-Sp6
C 28	84.4	3.5	542	13	CNS002YA	AL098284 Drosophila
C 29	84.4	3.5	865	13	AQ324474	AQ324474 m9x00018B
C 30	84.2	3.5	998	13	CNS00YXQ	AL096968 Drosophila
C 31	84	3.5	819	11	BG369277	BG369277 HVSMB1002
C 32	83.6	3.4	784	10	AL525973	AL525973 AL525973
C 33	83.6	3.4	907	13	CNS021J4	AL176953 Tetradon
C 34	83.6	3.4	942	13	CNS018GS	AL109318 Drosophila
C 35	83.2	3.4	717	13	CNS062NC	AL422542 T7 end of
C 36	83	3.4	870	13	AQ330286	AQ330286 nbx0046J
C 37	83	3.4	871	13	A2671726	A2671726 ENTHP05TF
C 38	82.8	3.4	843	13	A2551618	A2551618 ENTPV54TR
C 39	82.8	3.4	974	13	CNS001TT	AL075432 Drosophila
C 40	82.8	3.4	1147	13	B13042	B13042 T30W24-Sp6
C 41	82.6	3.4	849	13	A2546009	A2546009 ENTFW53TF
C 42	82.6	3.4	892	13	A2193174	A2193174 SP_1022.B
C 43	82.6	3.4	990	13	CNS006OI	AL065624 Drosophila
C 44	82.2	3.4	796	10	AL524807	AL524807 AL524807
C 45	82.2	3.4	835	13	A2188415	A2188415 SP_1011_B

ALIGNMENTS

RESULT 1	BF630719/C	standard; RNA; EST; 1337 BP.
ID	BF630719	
XX	BF630719;	
AC	XX	
SV	BF630719.1	
XX	XX	
DT	21-DEC-2000 (Rel. 66, Created)	
DT	21-DEC-2000 (Rel. 66, Last updated, Version 1)	
XX	XX	
DE	HVSMB0013H16f Hordeum vulgare seedling shoot EST library HVCN0002	
DE	(Dehydration stress) Hordeum vulgare cDNA clone HVSMB0013H16f, mRNA	
DE	sequence.	
XX	XX	
XX	EST.	
OS	Hordeum vulgare (barley)	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;	
OC	Triticeae; Hordeum.	
XX	XX	
RN	[1]	
RP	1-1337	
RA	Wing R., Close T.J., Kleinhofs A., Wise R., Begum D., Frisch D., Yu Y.,	
RA	Anderson H., Dale J., Henry D., Kernodle S., Palmer M., Rambo T.,	
RA	Saski C., Schwartzbeck J., Simmons J., Choi D.W., Main D., Wood T.,	
RT	"Development of a genetically and physically anchored EST resource for	
RT	barley genomics";	
RL	unpublished.	
XX	XX	
XX	Contact: Wing RA	
CC	Clemson University Genomics Institute	
CC	Clemson University	
CC	100 Jordan Hall, Clemson, SC 29634, USA	

CC	Tel: 864 656 7288
CC	Fax: 864 656 4293
CC	Email: rwling@clemons.edu
CC	Seq primer: ATTATACCTCCTACTAAGG
CC	High quality sequence start: 38
CC	High quality sequence stop: 1204.
XX	
FH	Key
FH	Location/Qualifiers
FT	source
FT	1. .1337
FT	/db_xref="taxon:4513"
FT	/note="Vector: lambdaZAP, Site_1: EcoRI, Site_2: Xho1"
FT	/organism="Hordeum vulgare"
FT	/cultivar="Morex"
FT	/clone="HVSMEB0013H16f"
FT	/clone_lib="Hordeum vulgare seedling shoot EST library
FT	HVCDNA0002 (Dehydration stress)"
FT	/tissue_type="Seedling shoot"
FT	/lab_host="HVC121"
XX	
SO	Sequence 1337 BP; 292 A; 22 C; 74 G; 941 T; 8 other;

Query Match	4.2%;	Score 103.2;	DB 5;	Length 1337;
Best Local Similarity	46.1%;	Pred. No. 3.2e-06;		
Matches 530; Conservative	0;	Mismatches 602;	Indels 18;	Gaps 5

Oy	821	tttttcaacgagtaatttaacgaacgaagataaaltctaccccttggaagaacttaag	880
Db	1333	TAAATTAAAAAAATATCTCTAAAAATTAATAATTAATTAATGAATATATAAAAATTA	1274
Oy	881	atcaacgagatgcgtcacaagatagtaaaaaatgggaaaaagataaaca---gcactacac	937
Db	1273	ATTAATATTAATAATATATTAATAATTAATAATTAATAATAATAATCTAAATAAAAA	1214
Oy	938	actggagcgatcttcattcatctgaagaagaagagactttaaagaagctgcagatcccta	997
Db	1213	ACAAATATCAATTAATATTAATAATTAATAATTAATAATTAATATTAATTAATAATCA	1154
Oy	998	ttagcgcaagaagaatgacgaatcattccttctcccaagaagaagaagctctca	1057
Db	1153	CAATTAATAATCTTTATTAATAATATATTAATTAATAATTAATAATTAATAATATATA	1094
Oy	1058	aaagaatacaatctgaatagtagatcttctca-----ctgaggaagaagct	1109
Db	1093	ATTAATATCAAAATATATAGTAATTAATTAATAATTAATTAATAATTAATAATTAATA	1034
Oy	1110	ttcaaaaaagctcaaatctgacatctcgatctccttaccgagaagaagaagactt	1169
Db	1033	AAATTAATATATAAAAAATTAATAATTAATTAATTAATTAATAATTAATAATTAATA	974
Oy	1170	aaatagaatcaggtgagtagtagtaaccccttaccgaaaaagaagaagcttctaa	1229
Db	973	TAAATACAAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATA	914
Oy	1230	aaagctgaaactgcatctcaacacatagatacttaacaaggtgcagaatacaggag	1289
Db	913	TAAATTAATAATTAATAAAGCAAAAAACAAAAACATGATTAAGTTTAATCATTAATA	854
Oy	1290	gttaattgtagtcgccaattcaatccttgatgtaagaagaagatataaaaggaattca	1349
Db	853	AAATATTAATGAATATTAATAATTAATAAT---AAATATTAATAATAATATACACACAT	796
Oy	1350	aaatatctgactcttaacatcaatccctggaagtaaccttgacataaattctt	1409
Db	795	AAATTAATAATTAATAATTAATTTTAATTAATTAATAATTAATAATTAATAATTAATA	736
Oy	1410	gtaagaataatgaatcaatcaatcaaccccttagctgcggaattagctgctc	1469
Db	735	AAAAAAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATA	676
Oy	1470	caactgaataactaaatcaatagsgtacttccaatgaattcaaaaaaattccaata	1529

D6	675	AAAAATTAATAAATAAAAAAAAAAAAT--AAAAAAATVAAAAAAATTAATAAAA	618
OY	1530	tagattcttcgttaactataatgatggtgataataagaaagccctcatagataatga	1589
D6	617	AATAATTAATAAAAAAAAAAAATTAATAAATAACAAAAATTAATAAATAAATAAAT	558
OY	1590	gcgttccaatcggagaacccaattaccacaagabactcgagcaagatatattagaanaatgg	1649
D6	557	AAATCAGATTCAAAATTAATAAATTAATAATCAAATTAATGATAAATAAATAAATAAAT	498
OY	1650	aaagcttatattaccaagaacatcgcgttcygaataaaggatgylacacaataatlaagca	1709
D6	497	AAACATTAACAACAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA	438
OY	1710	atccgaaaaagaatataaagatttgatgagcaagatgctgccaaagatgaataatagatcac	1769
D6	437	AAAAAAAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	378
OY	1770	aaaaatccaagaagcacagcttaaatataatccaagaatcgagataaagcataaggttaccc	1829
D6	377	AAATTAATAAATTAATAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA	321
OY	1830	aaaataccaaagcttatattacatccaagctgycataatagatatgcatccaatattgtaga	1889
D6	330	AAAAAATTAATAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	261
OY	1890	aagtcttatattaatgatgatgatgaaatggaataatcatccaagtgatctctataaaaaa	1949
D6	260	AATATTTAATTTTTTATTTAATTAATAAATAAATAAATAAATAAATAAATAAATAA	201
OY	1950	qgttaacaaat 1959	
D6	200	CGAATCGATT 191	
RESULT 2	CNS014J2/c		
LOCUS	CNS014J2	1025 bp DNA GSS	26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC BACN11L1 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL104216		
VERSION	AL104216.1	GI:5615827	
KEYWORDS	GSS.		
SOURCE	fruit fly.		
ORGANISM	Plasmid Drosophila melanogaster		
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1025)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)		
COMMENT .	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros.BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector peloBAC11.		
FEATURES	Location/Qualifiers		
source	1..1025		
	/organism="Drosophila melanogaster"		
	/plasmid="peloBAC11"		
	/db_xref="taxon:7227"		
	/clone_lib="DrosBAC"		
	/clone="BACN11L1"		
	/note="end : T7"		
BASE COUNT	148 a	88 c	6 g 522 t 261 others
ORIGIN			

BASE COUNT	148 a	88 c	6 g	522 t	261 others
ORIGIN					

Query Match	4.2%	Score 102.6	DB 13	Length 1025
Best Local Similarity	34.0%	Pred. No. 4.2e-06		
Matches 338	Conservative 102	Mismatches 553	Indels 2	Gaps 1
OY	959	aagaagaagagactttaaaagctgcagatctctatcttgcccaagaagaatgaca	1018	
DB	1008	AAAAAAAAAADAADAAAAAAAGWGTAAATKTAANATRTANCTKKTATGTTT	949	
OY	1019	taattcatcttctatctcaagaagaagctctcaaaagaatacaattgatatga	1078	
DB	948	KTTTTTMAAGWATGTCRATGVMAGAAAGCTTTCTDAAAATGKRWAMATGCTGAKTRT	889	
OY	1079	gtgatttttactactagtgaaagaagttttaaaaaagtcacaaatgatattccg	1138	
DB	888	TGGGDAARADAAAAGGTGATWAGRMAAATTTKMTKRTTAARTGCTGATARMAAAAA	829	
OY	1139	attcttacttgaaagaagaagactttaataagatacagctgagtagtagtaac	1198	
DB	828	AAAAAAAGWAGARKGAGAAAAATGCGTGMRAAAAAAAAKAAAAAGKGGAAAAAKR	769	
OY	1199	ctttactcgaagaagaagaagttttaaaaaagctgaaacttgattcaaccatag	1258	
DB	768	TTRTRGTGTAAAAAATATGTRTARWATTAATTTKTKTGAATGRTAARAWMAAA	709	
OY	1259	atatatacaaggtgcaagatcacagagggttaattgattagtcgctcaattaatc	1318	
DB	708	ATRTAAATWATGAAKTGRMAAAAAAARAAAAATGTRMAWATTAATTTKATGTGA	649	
OY	1319	atgtaagaagacagataaaaggatattcaaatatgtagcttattatcaacaatca	1378	
DB	648	ATWKGCTKCTKKMAAAAAAAMAAAKWRTAAWKRTAARAAAMDGTCKTKTKTTTRA	589	
OY	1379	ttggaagtaacctgtcaacaatttacttgtaagaatatgaatcatcaacttca	1438	
DB	588	AM-AAKGRGTGCKRAXKTTTRRDWMAAAAAAAGGCGTGAADARA	531	
OY	1439	caggacacctagtgcgatttagttgattccactgataataataataagagga	1498	
DB	530	AKGDAAARADAGRWMAAAAKRTATKGTGKTKTGTGTAAGAKTAAATWMAAAG	471	
OY	1499	tttcaatgattcaaaaaaaattcaaatatagtaattcagtaactatgatctgtg	1558	
DB	470	TGTGAKKMTATTTAAAAAATAAAAKKRTGRTGRCRAATTTATRTAKKTTAA	411	
OY	1559	atataaagaaagcctgcattagatacgagcttgaatgaggaatccaattcac	1618	
DB	410	AAAAAATGTRKATRTKTKTGCDKAAARAADAAATTTAAAAAATAAAAAA	351	
OY	1619	cagatctcgagcagatatttagaaaatgaaaagcttatatacaagaacaatcg	1678	
DB	350	AAAAAANANAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	291	
OY	1679	tggaaataaagagatgcaataatlaagcaatccgaagaagatatataaggtgatg	1738	
DB	290	AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	231	
OY	1739	cgaaagtagtgcgaagagtaaaatagatacaaaaatccaagaagcaggttaataa	1798	
DB	230	AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	171	
OY	1799	atcaggaatcggaataaagattaggttccaaatatatacaagcttatataactca	1858	
DB	170	NNNNNNNNNANNNANNNANNNANNNANNNANNNANNNANNNANNNANNN	111	
OY	1859	tgcataatagatagcacaatattgtagaagtgcttattcaattgaagcaatgga	1918	
DB	110	CMAATCTTTTATTAATTTNNNNNNNANANANANANANNNNNNNNNNNNN	51	
OY	1919	aaaaataatctcaagtgatcttataaaaaagga	1953	
DB	50	NNNNNNNANNN	16	

RESULT	3			
CNS0021J				
LOCUS				
DEFINITION	CNS0021J	1101 bp	DNA	GSS
ACCESSION	Drosophila melanogaster genome survey sequence TET3 end of BAC #			
VERSION	BACR05N11 of RPCI-98 library from Drosophila melanogaster (fruit			
KEYWORDS	fly), genomic survey sequence.			
SOURCE	AL061936			
ORGANISM	AL061936.1	GI:4940214		
REFERENCE	GSS.			
AUTHORS	fruit fly.			
TITLE	Drosophila melanogaster			
JOURNAL	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
COMMENT	pteriygota; Neoptera; Endopterygota; Diptera; Brachycera;			
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
	1 (bases 1 to 1101)			
	Genoscope.			
	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :			
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr			
	- Web : www.genoscope.cns.fr)			
	Determination of this BAC-end sequence was carried out as part of a			
	collaboration with the Berkeley Drosophila Genome Project (BDGP).			
	The BDGP is constructing a physical map of the Drosophila			
	melanogaster genome using these BACs. For further information			
	please see http://www.fruitfly.org The BDGP Drosophila			
	melanogaster BAC library was prepared by Kazutyo Osoegawa and			
	Aaron Mammeter in Pieher de Jong's laboratory in the Department of			
	Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,			
	NY. The library is named RPCI-98 and was constructed by partial			
	EcoRI digestion of Drosophila DNA provided by the BDGP from the			
	isogenic strain y ² . cn bw sp. the same strain used for the BDGP's			
	P1 and EST libraries. A more detailed description of the library			
	and how to order individual BAC clones, the entire library, or			
	filters for hybridization from the BACPAC Resource Center can be			
	found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .			
FEATURES	Location/Qualifiers			
source	1..1101			
	/organism="Drosophila melanogaster"			
	/db_xref="taxon:7227"			
	/clone_lib="RPCI-98"			
	/clone="BACR05N11"			
	/note="end : TER3"			
BASE COUNT	631 a 7 c 28 g 289 t 146 others			
ORIGIN				
Query Match	3.9%	Score 93.8	DB 13	Length 1101
Best Local Similarity	40.2%	Pred. No. 8e-05		
Matches 346	Conservative 54	Mismatches 459	Indels 2	Gaps 1
OY	134	aagaagaagaataaataagatgagataagaagaatgaagaacgaataaacac	193	
DB	199	AA	258	
OY	194	aggaagagccttaagaatcatgaacacattgaaataagaaagtaagggagag	253	
DB	259	AA	318	
OY	254	aagctgttaaaagaagcgacgaagaagctacttggaagatccatcgatgtttag	313	
DB	319	AA	378	
OY	314	agatgataaagcaattggaagaagatatactgtgga tgg tgaatatatacaacata	373	
DB	379	AA	438	
OY	374	tatcttagaagcattatctgaagataagaataaataaagcattatcggaagatg	433	
DB	439	AA	498	
OY	434	cttattacatgacattatgatatgcaagaagagatagaaacccgactctgtaatcc	493	

Db 499 AAAAAAAAAAATAATTAATTTTWTATTAATTAATTTTWTATTTT 558
QY 494 aatctcggaagatcagtagaanaatcggaaagcactgaacgtttatcagaatag 553
Db 559 WTTAATTTTAAATTTTAAWMAAATTTTAAATAAAMATTTTAAATTTTAAWMAA 618
QY 554 gtaagatctatcgaagagatatttaagtaaaatcaatcaacatatacagaattttag 613
Db 619 AAAAAATTTTAAAMATTTTWTATTAATTAATAAATAAATAAATAAATAAATAAATAA 678
QY 614 atgctataaataccatcaaaatgcatcga--ttcagatgagcaagatcctttatcac 671
Db 679 WMAATTAATTAATTTTAAATAAATTTTAAATTTTAAATAAATAAATAAATAAATAA 738
QY 672 taatcagctaaaggaacatccacagactttctgtagaattctggaacaatagcaa 731
Db 739 AAAAATATATATTTTAAATTTTAAATTAATAAATAAATAAATAAATAAATAAATAA 798
QY 732 tgaggtcacagaagatattcggaaagctttgcataatatacgcacagcatctga 791
Db 799 TTAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 858
QY 792 tgtttacagcttatgcaccggaagctttaattacatgagataaatttaacggaacaga 851
Db 859 TAAATTTATTTTATTTTAAATAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 918
QY 852 aataatctatccttggaaagactaaagatcaacgagatgctgcaagatggaanaatg 911
Db 919 AATGKAATGKAKKGGTAAARAATGTAAGAARATTAATGTTTAAATGTTTAAATGTTTAA 978
QY 912 ggaagaagataaagcactcaacactggaagcagatccttatcagaagaagaagg 971
Db 979 AAAAAAARAKGAAAAATGAARAGATGAATAATTTTAAATAAATAAATAAATAAATAA 1038
QY 972 actttaaaaaagctgcagat 992
Db 1039 AAAAAAAAAAAGAGAKAK 1059

RESULT 4
LOCUS CNS0022B/c 1059 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACN01115 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL097133 GI:5608744
VERSION AL097133
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Braachyera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1059)
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaut at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES
source
1. 1059
Location/Qualifiers
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"

Query Match 3.8%; Score 92; DB 13; Length 1059;
Best Local Similarity 37.5%; Pred. No. 0.00015;
Matches 348; Conservative 87; Mismatches 491; Indels 3; Gaps 1;
BASE COUNT 14 a 21 c 66 g 760 t 198 others
ORIGIN
/clone_11b="DrosBAC"
/clone="BACN01115"
/note="end : T7"
QY 1033 tctcaagaagaagaagagcttctaagaagaatcacaaattgtagtagatctttact 1092
Db 972 KCTCTWCAABAABAAACMBCKBCKGHCYBCKBCKWMAAADRADCKYKBABBAAD 913
QY 1093 actgagaaagaagagcttttaaaagctacaatgtagatcgtgacttcttaatcga 1152
Db 912 AAAAAAACAACAGNMAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 853
QY 1153 gaagaagaagagctttaaataagatagacaggtgtagtagtaatccttaatcga 1212
Db 852 AAAAAAACAACAGNMAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 793
QY 1213 gaaaagagcttttaaaagctgaactgtagatattacacatatgataatca---a 1269
Db 792 AMACBCKACACBCKAC 733
QY 1270 aggtgcagaatacagaagggttaattgtagtcgcgtcaattcctgtagtaagaag 1329
Db 732 AACSCSKABABCKCAAAAKMAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 673
QY 1330 cagatataaaggagatcaataatgtagtcttattatcaatccatctggaatcc 1389
Db 672 AAARKAAADADGAKAKAAATDKKKKKAAATGRDKKKADDAKAKAAAAA 613
QY 1390 ttgtacaataaattatttcttgaagaataatgataatcaatccatccagcaacct 1449
Db 612 DRGABAAAKKAAKAGDAKAGGKGTGAGTKRGDAAAAAATAAATAAATAAATAAATAA 553
QY 1450 ggtcggaattgattgattccactgatatactaaatgaatgaaggtatttcaatga 1509
Db 552 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 493
QY 1510 ttcaaaaaaattcaaatatagtagtcttctgtaacatatgattgtagtataagaa 1569
Db 492 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 433
QY 1570 aggcctgcatagtagtagagcgtttggaatgggaatccaatatccacgatactga 1629
Db 432 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 373
QY 1630 gcagatatattgaagaatggaagcttataatcaagaagaacatcgctcgtgaataag 1689
Db 372 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 313
QY 1690 gatgtacaataatgaagcaatccgaagaagaatatataagattgtagcgaagtg 1749
Db 312 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 253
QY 1750 ccaagaagtagtaagtagtacaagaatccagaagcagcgttaaatataatccaggaatg 1809
Db 252 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 193
QY 1810 aataagaatcaggttaccacaaatatacaagaagcttattacattcaatcgtcataatga 1869
Db 192 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 133
QY 1870 tatgcataccaatatgtagaagtgcttatttaatatgaatgaatgaagaataatatt 1929
Db 132 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 73
QY 1930 caaagtcatctataaagaagtagtaacaa 1958

collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genomics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source
1..1036
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_1lb="RPCI-98"
/clone="BACR11116"
/note="end : TET3"
BASE COUNT 64 a 56 c 41 g 701 t 174 others
ORIGIN

Query Match 3.8%; Score 91.6; DB 13; Length 1036;
Best Local Similarity 39.3%; Pred. No. 0.00017;
Matches 305; Conservative 65; Mismatches 405; Indels 1; Gaps 1;

Oy 1040 aagaaagagcttctaaagaatacaatgatagtagtcttcttctactcagag 1099
Db 797 AAAAAAATCCAGHACAYAAAAAAYGASTCTAAATAGAAAAATATCCSMS 738
Oy 1100 aaaaagagcttctaaagaatacaatgatagtagtcttcttctactcagag 1159
Db 737 AAHAHAACCTAGAAAAAARAGAAVAAAAAGYHBAKTKCKCAAMAMVCTTAAT 678
Oy 1160 aagagcttctaaagaatacaatgatagtagtcttcttctactcagaaagaa 1219
Db 677 AABKAKCTTAAAGAAVAAAAVMAARCAAAAAATGISTTACATRAAAAAARRAA 618
Oy 1220 agcttctaaagaatacaatgatagtagtcttcttctactcagag 1279
Db 617 AARVMAAAAAAATAATATCGAAMAGAAAAAGACTAATATAGYATGAKAMAA 558
Oy 1280 aacagagaggttaatgatagtagtcttcttctactcagagagagagagag 1339
Db 557 TAAAKAAAGTSTRAATATWAKAAAGAAANAKAGRTDGAKAAAAAATRTADAA 498
Oy 1340 gggataatcaaatatgcttcttctactcaatcacttctgagagagagagag 1399
Db 497 AAAAAAATAAATG-KBTTTBAAAAAARAGADAMWAKAGDGGGAAKCKAKAWATA 439
Oy 1400 aaattatttgaatgaataatgaataatgaataatgaataatgaataatgaata 1459
Db 438 AATATTAATCTATTAATAAAGATKKKARAGAKAAAAAANANANANANANANAN 379
Oy 1460 tagtgatctccactgataactaataatgaagtagtcttctactcagagagag 1519
Db 378 AAAAAAATAAATG-KBTTTBAAAAAARAGADAMWAKAGDGGGAAKCKAKAWATA 319
Oy 1520 attcaaatatgtagtcttctactgataatgtagtcttctactcagagagagag 1579
Db 318 AAAAAAATAAATG-KBTTTBAAAAAARAGADAMWAKAGDGGGAAKCKAKAWATA 259
Oy 1580 tagaataagagcgttgaatgaatgaatgaatgaatgaatgaatgaatgaatga 1639
Db 258 AAAAAAATAAATG-KBTTTBAAAAAARAGADAMWAKAGDGGGAAKCKAKAWATA 199
Oy 1640 tagaataagagcgttgaatgaatgaatgaatgaatgaatgaatgaatgaatga 1699
Db 198 AAAAAAATAAATG-KBTTTBAAAAAARAGADAMWAKAGDGGGAAKCKAKAWATA 139

Oy 1700 taattagcaatccgaagaataatgaatgaatgaatgaatgaatgaatgaatga 1759
Db 138 AAAAAAATAAATG-KBTTTBAAAAAARAGADAMWAKAGDGGGAAKCKAKAWATA 79
Oy 1760 aattagatcaaatatcaagaagcagtagtcttctactcagagagagagagag 1815
Db 78 AAAAAAATAAATG-KBTTTBAAAAAARAGADAMWAKAGDGGGAAKCKAKAWATA 23

RESULT 7

BE420745

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..1885

/organism="Hordeum vulgare"

/cultivar="Barke"

/db_xref="taxon:4513"

/clone="HMM002.B02"

/clone_1lb="ITPC HMM Barley Leaf Library"

/tissue="leaf"

/dev_stage="14 day old"

/note="Vector: pBluescriptSK(-); 850 bp average insert size."

BASE COUNT 1138 a 219 c 212 g 176 t 140 others

ORIGIN

Query Match 3.7%; Score 90.8; DB 10; Length 1885;
Best Local Similarity 38.8%; Pred. No. 0.00015;
Matches 386; Conservative 0; Mismatches 610; Indels 0; Gaps 0;

Oy 959 aagaagaagagcttctaaagaataatgaatgaatgaatgaatgaatgaatga 1018
Db 890 AAAAAAATAAATG-KBTTTBAAAAAARAGADAMWAKAGDGGGAAKCKAKAWATA 949
Oy 1019 taattcattcttctctcagaagaagaagcttctaaagaataatgaatgaatga 1078
Db 950 NNN 1009
Oy 1079 gtagcttcttctctcagaagaagaagcttctaaagaataatgaatgaatga 1138
Db 1010 AAAAAAATAAATG-KBTTTBAAAAAARAGADAMWAKAGDGGGAAKCKAKAWATA 1069
Oy 1139 attcttctcagaagaagaagcttctaaagaataatgaatgaatgaatgaatga 1198
Db 1070 AAAAAAATAAATG-KBTTTBAAAAAARAGADAMWAKAGDGGGAAKCKAKAWATA 1129

[illegible]

COMMENT	Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .	
FEATURES	Location/Qualifiers
SOURCE	1. 1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_11b="RPCI-98" /clone="BACR29B23" /note="end : 17"
BASE COUNT	419 a 91 c 60 g 299 t 232 others
ORIGIN	
Query Match	3.7%: Score 89.8; DB 13; Length 1101;
Best Local Similarity	36.7%: Pred. NO. 0.00031;
Matches 232; Conservative 112; Mismatches 274; Indels 15; Gaps 3.	
QY 1348 caaaatcgtgccttaccatcaatccatcggagctgtgacataaatlta	1407
Db 469 MWMATWTWMAAAMAAATTTATWATWMAAAMAMMMWATTTTMMWMTATTTTWA	528
QY 1408 ttgctgaaaaatagaatatacaatccctcagcaacctgagcgagatttgat	1467
Db 529 WMTTATWMTAAAMAAAMAAATTTTAAAMAAATWATTAAMATTTTAAAMATTTAA	588
QY 1468 tccactgtaatacctcaaaatlaatagaggtattcctcaagaaatcctcaaaaaa	1520
Db 589 TWTATTAATTTWATTAATTAATTAATAAATATTTTWTWATAATTTTAAATTAATTA	648
QY 1521 ttccaatagatattcctagtaacctataatgatttgatataaatgaagccgcatt	1580
Db 649 WTTATTTAATAATWATTAATTTWMTAAATTTAAATTTWMAAAMMAAAMMAAMMT	708
QY 1581 agataatgacgttggaaatggagatcccaatcaaccag-atactgagcagatatt	1639
Db 709 WAAWATATATATATTAATTAATAAAMAAATTAAMAAAMAAATATATATATATAT	768
QY 1640 tagaaatcgaagccttatatacaagaagacacgcgtctggaataaagaatgacaa	1699
Db 769 TWTAAWMAATWMAAMWMTATATWATMAATATWMAAMWMTAAATWATWMAAAMWA	828
QY 1700 ta-----attaagcaatccgaagaaagatatlaagatgatgcgaagtagcca	1752
Db 829 TAWMATWATATWMAATWMAAMAAATWATATATWATAATWATAAAMAAATATWMTW	888
QY 1753 aagaataatagatacaaaaatcgaagaacagcagtaataataatcaagaatcggaat	1812
Db 889 TWTTTTWWMAAMWATTAAMAAAMWATTAAMAAAMMAAAMMAAAMMAAMWMTATTTWT	948
QY 1813 aaagcattagaggtaccaaaatacaagaagccttatatacattcaacgtgataatgat	1872
Db 949 ATTAAATWATATWMAATWATTAATWMAATWATATATTTATATATTTWATWMAATATAT	1008
QY 1873 gcaaccatatgtgaaagcgtcattataatcgaatgaatgaaagaaataatattcaa	1932
Db 1009 TTATTAAMWATATTTTAAAMWATATATATATWMTAMWMTATATAMMAAATTTAMTTAT	1068
QY 1933 agtgcataaaaaagtgtaacaattactta 1965	

[illegible]


```

FH source 1. 838
FT /db_xref="taxon:4513"
FT /note="Vector: LambdaZAP; Site_1: EcoRI; Site_2: XhoI"
FT /organism="Hordeum vulgare"
FT /cultivar="Morex"
FT /clone="HVSMEB0010613f"
FT /clone_lib="Hordeum vulgare seedling shoot EST library
FT HYCDNA0002 (Dehydration stress)"
FT /issue_type="Seedling shoot"
FT /lab_host="TJC121"
XX
SQ Sequence 838 bp; 268 A; 19 C; 58 G; 492 T; 1 other;

```

```

Query Match 3.6%; Score 86.8; DB 5; Length 838;
Best Local Similarity 47.8%; Pred. No. 0.00092;
Matches 357; Conservative 0; Mismatches 373; Indels 17; Gaps 3;

```

```

OY 1116 aaagctcaaatgatgatcgtgattcttctatctgaagaagaagagctttaaag 1175
    || || || || || || || || || || || || || || || || || || || ||
DB 789 AATTATAGATTATTAAGCCGTTATTTATCTGCTTAATATAAATTAATATAAAT 730
OY 1176 aatacaggtgagtagtagtaatccttctatctgaagaagaagagctttaaag 1235
    || || || || || || || || || || || || || || || || || || || ||
DB 729 AAAATTAATTAATAAAGTACCAATTAATATAAATAAATAAATTAATGATACATAT 670
OY 1236 gaacttgatataccaacatagatatttaacaaaggtcgaagatagagaggtta 1295
    || || || || || || || || || || || || || || || || || || || ||
DB 669 AATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 610
OY 1296 -----tgatagtcgtcaatcttgatgtaagaagcagtlataaaggagat 1348
    || || || || || || || || || || || || || || || || || || || ||
DB 609 AATATATTTTAACTCGATTAAATAAATAATTAATTAATTAATTAATTAATTAAT 550
OY 1349 aaaaatctgagcttctatctatccatctggaagcttctgtaacaataaattat 1408
    || || || || || || || || || || || || || || || || || || || ||
DB 549 TATATATATTTTAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 490
OY 1409 tgaatgaataatgaataatacaaccctcagcaagccaggtcgagattgagtc 1468
    || || || || || || || || || || || || || || || || || || || ||
DB 489 TATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 430
OY 1469 ccactgataactaaatagaggtatttcaatgagatcaa--aaaaatttcaa 1526
    || || || || || || || || || || || || || || || || || || || ||
DB 429 ATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 370
OY 1527 atatagtaattctgaataatataatgttgttgaataatgaagagctgcatagaa 1586
    || || || || || || || || || || || || || || || || || || || ||
DB 369 TTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 310
OY 1587 tggagcttggaatggaagatccaattatcaacagatactcgagcagagatattaga 1646
    || || || || || || || || || || || || || || || || || || || ||
DB 309 AAAAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 250
OY 1647 tggaa-----agcttatattcaagaagaacatcgctcggaaataaagggtgta 1698
    || || || || || || || || || || || || || || || || || || || ||
DB 249 AAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 190
OY 1699 ataataagcaatcggaagaatataatagattgagtggaagtgagtcgaagaag 1758
    || || || || || || || || || || || || || || || || || || || ||
DB 189 TTTAATTAATTTCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 130
OY 1759 aaaaatgatacaaaaatcgaagaagcaggttaataatacaaggaatggaataag 1818
    || || || || || || || || || || || || || || || || || || || ||
DB 129 AAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 70
OY 1819 ctagggttaccacaatatacaagaagctt 1845
    || || || || || || || || || || || || || || || || || || || ||
DB 69 TCATTTTAAATTTTTTTAAATAAATCACT 43

```

RESULT 14

```

CNS00A4L
LOCUS CNS00A4L 963 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR20A24 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL054918
VERSION AL054918.1 GI:4935889
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 963)
REFERENCE
Genoscope.
Direct Submmission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammeter in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
Location/Qualifiers
source 1. 963
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR20A24"
/note="end : TET3"

```

```

BASE COUNT 504 a 25 c 41 g 128 t 265 others
ORIGIN

```

```

Query Match 3.6%; Score 86.8; DB 13; Length 963;
Best Local Similarity 41.4%; Pred. No. 0.00088;
Matches 208; Conservative 55; Mismatches 237; Indels 3; Gaps 1;

```

```

OY 132 aaaaagagaagagaataaagtgagaaatgaagaagaatgaagaagaataaag 191
    || || || || || || || || || || || || || || || || || || || ||
DB 344 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 403
OY 192 acaggaagagcatctaaaggaatcatgaacacatctgtaaaatgaagtaaaaggga 251
    || || || || || || || || || || || || || || || || || || || ||
DB 404 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 463
OY 252 ggaagctgtctaaagaagcagcagaagctacttgagaaggtaccatctgattc 311
    || || || || || || || || || || || || || || || || || || || ||
DB 464 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 523
OY 312 agagatgtataagcaatctggaggaagatatatctgtgagtgagatcttaagaa 371
    || || || || || || || || || || || || || || || || || || || ||
DB 524 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 583
OY 372 tatactctagaagcatctatctgagaataagaataaataaagacattatggaaga 431
    || || || || || || || || || || || || || || || || || || || ||
DB 584 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 643
OY 432 tgccttatacatga---acattatgataatgaagaagatatgaacgcgactgtc 488
    || || || || || || || || || || || || || || || || || || || ||
DB 644 WTATTTAKAAKAATTAATAAAGAARAKTAAGAAAGDAATAATADTFRKKAATAAATKKK 703

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 2, 2001, 13:49:05 ; Search time 54.73 Seconds
(without alignments)
10055.573 Million cell updates/sec

Title: US-09-747-521-1
Perfect score: 2430
Sequence: 1 atgaataataaaaaagaatt.....agttcatattactactataa 2430

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/1na/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/1na/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/1na/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/1na/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2430	100.0	3291	1	US-08-021-601-1	Sequence 1, Appli
2	2430	100.0	3291	1	US-08-082-849B-1	Sequence 1, Appli
3	2430	100.0	3291	5	PCT-US94-01624-1	Sequence 1, Appli
4	762.4	31.4	1368	1	US-08-021-601-5	Sequence 5, Appli
5	762.4	31.4	1368	1	US-08-082-849B-5	Sequence 5, Appli
6	762.4	31.4	1368	5	PCT-US94-01624-5	Sequence 5, Appli
7	762.4	31.4	1425	1	US-08-021-601-7	Sequence 7, Appli
8	762.4	31.4	1425	1	US-08-082-849B-7	Sequence 7, Appli
9	762.4	31.4	1425	5	PCT-US94-01624-7	Sequence 7, Appli
10	762.4	31.4	1524	1	US-08-021-601-9	Sequence 9, Appli
11	762.4	31.4	1524	1	US-08-082-849B-9	Sequence 9, Appli
12	762.4	31.4	1524	5	PCT-US94-01624-9	Sequence 9, Appli
c 13	91.6	3.8	7218	1	US-08-232-463-14	Sequence 14, Appli
14	66.4	2.7	5361	4	US-08-973-462-2	Sequence 2, Appli
15	66.4	2.7	5361	4	US-08-973-462-1	Sequence 2, Appli
16	66	2.7	4766	5	PCT-US93-07261-10	Sequence 10, Appli
17	65	2.7	3095	6	5231168-1	Sequence 10, Appli
18	65	2.7	19124	2	US-08-487-826B-13	Patent No. 5231168
19	64.2	2.6	8920	2	US-08-446-855A-1	Sequence 13, Appli
20	64.2	2.6	8920	4	US-09-150-741-1	Sequence 1, Appli
c 21	61.4	2.5	615	4	US-08-998-416-186	Sequence 1, Appli
22	61.2	2.5	696	4	US-09-461-697-193	Sequence 186, App
23	61.2	2.5	699	4	US-09-461-697-191	Sequence 193, App
24	61.2	2.5	717	4	US-09-461-697-189	Sequence 191, App
25	61.2	2.5	774	4	US-09-461-697-187	Sequence 189, App
26	61.2	2.5	819	4	US-09-461-697-185	Sequence 187, App
27	61.2	2.5	1669	4	US-09-461-697-184	Sequence 185, App
						Sequence 184, App

28	59.2	2.4	2223	1	US-08-257-073-4	Sequence 4, Appli	
29	58	2.4	5852	1	US-07-867-106-2	Sequence 2, Appli	
30	57.2	2.4	5433	3	US-08-929-329-1	Sequence 1, Appli	
c	31	57.2	2.4	19124	2	US-08-487-826B-13	Sequence 13, Appli
c	32	56.6	2.3	837	4	US-08-998-416-288	Sequence 288, App
c	33	56.2	2.3	5852	1	US-07-867-106-2	Sequence 2, Appli
c	34	55.4	2.3	6243	2	US-09-056-075-1	Sequence 1, Appli
c	35	53.6	2.2	636	4	US-08-998-416-1137	Sequence 1137, Appli
36	52.2	2.2	3337	1	US-08-072-610-1	Sequence 1, Appli	
37	52.2	2.1	3337	2	US-08-719-822B-1	Sequence 1, Appli	
38	52.2	2.1	3337	4	US-09-092-458-1	Sequence 1, Appli	
c	39	51.8	2.1	658	4	US-08-998-416-595	Sequence 595, App
40	51	2.1	636	4	US-08-998-416-1137	Sequence 1137, App	
41	50.6	2.1	1654	3	US-08-913-842-20	Sequence 20, Appli	
42	50.6	2.1	1744	3	US-08-913-842-27	Sequence 27, Appli	
43	50.6	2.1	1875	3	US-08-913-842-1	Sequence 1, Appli	
44	50.6	2.1	1965	3	US-08-913-842-2	Sequence 2, Appli	
45	50.4	2.1	663	4	US-08-998-416-191	Sequence 191, Appli	

ALIGNMENTS

RESULT 1

US-08-021-601-1

; Sequence 1, Application US/08021601

; Patent No. 5591631

; GENERAL INFORMATION:

; APPLICANT: Leppla, Stephen H.

; APPLICANT: Klimpel, Kurt R.

; APPLICANT: Nichols, Peter J.

; APPLICANT: Atora, Naveen

; APPLICANT: Singh, Yogendra

; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESS: Needle & Rosenberg, P.C.

; STREET: 133 Carnegie Way, Suite 400

; CITY: Atlanta

; STATE: Georgia

; COUNTRY: USA

; ZIP: 30303

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/021,601

; FILING DATE: 19930212

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Spratt, Gwendolyn D.

; REGISTRATION NUMBER: 36,016

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 404/688-9880

; TELEFAX: 404/688-9880

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3291 base pairs

; TYPE: NUCLEIC ACID

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; ORIGINAL SOURCE:

; ORGANISM: Bacillus anthracis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 580..2907

; US-08-021-601-1

Query Match 100.0%; Score 2430; DB 1; Length 3291;
Best Local Similarity 100.0%; Pred.No. 0;
Matches 2430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 atgaataaaaaagaattataaaagttaattagtatgtcaatgttttaagtaacagcaatt 60
|||||
481 ATGAATATTAATAAAGAAATTTATTAAGTAATGTATGTCATGTTAGTACAGCAATT 540
61 accttgagtggtcccgctcttacccttgtaacagggcgcggtgtcaatgtgtatgta 120
541 ACCTTGAGTGTCCCGCTTATCCCTTGTAACAGGGGGGGCGGTGATGATGATGTA 600
121 ggtatgacgttaaaagagaagaataaagaatgaataaagaataaagaatgaatga 180
601 GGTATGACGCTTAAGAAAGAAAGAAATAAATGATGATGAATAAGAAAGAAAGAA 660
181 cgaataaacaacagagaaggcatcttaaggaaatcgaacacatgttaaaaaatga 240
661 CGAATATTAACACAGAAAGCAATTTAAAGAAATCATGAACACATTTGTAATAATGAA 720
241 gtaaaaggaggaaagcgtgttaaaaaaaggcagcagaaaagctacttgaaagtaca 300
721 GTAAGAGGGAGAGCTGTTAAAAAGAGGACAGAAAAAGCTACTGAGAAAGTACCA 780
301 tctgagtctttagagatgtataaagcaattggagaaagatatatctgtgagtgtat 360
781 TCTGATGTTTGAAGATGTATAAGCAATTTGAGAGAAAGATATTTATTTGCGATGCTGAT 840
361 attaacaacatatcatctttagaagcatctcgaagataaagaataaataaagaatc 420
841 ATTAACAACATATATCTTTTGAAGCATTTATCTGAAGTAAAGAAAAATTAAGACATT 900
421 tatggagaagatgctcttactacatgaacatatagtatatacgaagaagatatga 480
901 TATGGAGAAAGATGCTTTATTACATGAACATTTATGATATGCAAAAGAGATGAAACCC 960
481 gtaactgtaatccaaactctcggaagatatagtagaataaatactgtaaaagcactga 540
961 GTAATTTTAAATCCAAATGTCGGAAGATTATGTAAATAATGTAAGAAAGGCACTGAACGTT 1020
541 tatatgaataagtaagatacttaagaagatatcttaagtaaatcttaacacatat 600
1021 TATATGAATAGGTAAAGATTTTCAAGGATTTTAAAGTAAATTAATCAACCATAT 1080
601 cagaatctttagatgtataaataacatataaataatgacatctgaatcgaatga 660
1081 CAGAAATTTTGAATGTATTAATAATACCATTAATAATGATCATGATGAGACAGAT 1140
661 ctttatattactaaatcgacttaagaaacatcccaacacatttctgtagaattcttgaa 720
1141 CTTTATTTACTAATACGCTTAAGAAACATCCACAGACTTTTGTAGAAATTTTGGA 1200
721 caaaatgacatgagtaacagaagatcttgcaaaagctttgcatatataatcgagcca 780
1201 CAAATATGCAATGAGTACAAAGAGTATTTGCGAAAGCTTTTGCATATTAATGAGACCA 1260
781 cagcatcgtgaatgttcttaagccttatgcacccggaagctttaatcatatgaataatc 840
1261 CAGCATGCTGATGTTTACAGCTTTATGACCCGGAAGCTTTAATTAATGATGAATAATTT 1320
841 aaggaacaaagaataatctcatctctgtgaagaacttaagaatcaagagatgctcaaga 900
1321 AACGAACAAAGAAATTAATCTATCTTGGAAGAACCTTAAGATCAACGATCTCTCAGA 1380
901 tatgaataatgggaagaataaacaacacatacaacacatgagacggaattcttcttga 960
1381 TATGAAAAATGGAAGATTAACACACATATCAACCTGAGCGATTTCTTTATCTGGA 1440
961 gaaggaagaggaactttaaagaagctgcagatctcctaattgagccaaagaagaatga 1020
1441 GAAGGAAGAGGACTTTTAAAAAGCTGAGATTCCTTATGTAGCCAAAGAAAGATGACATA 1500

1021 attcatctcttatctcaagaagaagaagccttctaagaagaataacaaatgtatagtagt 1080
|||||
1501 ATTCAATCTTATCTCAAGAAAGAAAAAGACCTTCTAAAAAGAAATACAAATGATAGTAGT 1560
1081 gatctttactctactgaggaagaagcttctaagaagaataacaaatgtatagtagt 1140
1561 GATTTTATCTACTGAGAAAAAGACTTTTAAAAAGCTACAAATGATATTCGCGAT 1620
1141 tcttatctgaagaagaagaagcctttaaataagaatcagatgtagtagtaatcct 1200
1621 TCTTATCTCAAGCAAGAAAAAGACCTTTAAATGAAATACAGGGGTATGATAGTAAATCCT 1680
1201 ttatctgaagaagaagaagctttaaagaagctgaagaactgtatatacacaatagat 1260
1681 TTATCTGAAGAAAGAAAGAGCTTTTAAAAAGCTGAACCTGATATTCACACATATGAT 1740
1261 ataatcaaaagttgcaagatcacagagggttaatatatagtcgcgtcaatctatgac 1320
1741 ATTAATCAAAAGTTGCAAGATACAGAGGGTTAATTAATGATGCTGTTATTAATCAATCCATT 1800
1321 gtaagaagcagtaataaaggatatacaaaataatgaatgcttatatacaatccat 1380
1801 GTAAGAAAGCAGTATAAAGGATATCAAAATATGATGCTGTTATTAATCAATCCATT 1860
1381 ggaagtaccttgtaacaataaatttatgtatagaataatgaataatcaatcaatcaca 1440
1861 GGAAGTACCTTGTAACAATAAATTTATTTGATGAAAAATATGAATATCAATCAATCCATTACA 1920
1441 gcaacccataggtgcggaattagttgatccactgataatcctaataatgaagaagatc 1500
1921 GCAACCTTAGTGCGGATTTTATGTTGATTCACATGATATTAATTAATTAAGAGTAT 1980
1501 ttcaatgaattcaaaaaaaatltcaaalatagtaattctagtaactatagatgtgtat 1560
1981 TTCAATGAATTCAAAAAAATTTCAAAATATGATTTCTGTAACATTAATGATGATGAT 2040
1561 ataaatgaaggcctgcatagataagatgaagcgttgaagaagggaatccaattatccca 1620
2041 ATTAATGAAGGCTGCTGATTAATGATTAATGAGCGTTTGAAAGGAGAAATTCATTAATCA 2100
1621 gatactgagcagagatacttagaataatgaaagccttatatacaagaagaacatcgctgt 1680
2101 GATCTGAGCAGAGATTTTGAAGAAATGGAAGCTTTATTAACAAAGAAACATCGCTG 2160
1681 gaaataaagatgtacaataaattaagcaatccgaagaagaataatagaatgtatgctg 1740
2161 GAAATTAAGGATGTACAAATTAATTAACCAATCCAAAAAAGAAATTAATGAAGATGATGCG 2220
1741 aaagttagtgcgaagaagtaaatagatcacaaaatltcaagaagaagcagttaaataat 1800
2221 AAAGTAGTGCCTCAAGAGATTAATAATGATTAACAAAAATTCAGAGAGCACAGTTAAATATTAAT 2280
1801 cagaatggaataaagcattaggtgtaccaaataacaaagccttatatacctcaagctg 1860
2281 CAGGAATGGAATAAAGCATTAAGGTTTACAAAAATATCAAAAGCTTATTAATCAAGCTG 2340
1861 cataatagatatagccaaatattgtagaagtgcttaataatcaatgaaatggaaa 1920
2341 CATATATGATATGATCAATATATGTAGAAAGTCTTATTAATTAATGAATGAGAAA 2400
1921 aataataatcaaaagtgtattataaaagaagtaacaaatcttaagtgtgtgaatgga 1980
2401 AATATATTTCAAAAGTGAATCTTATTAATAAAGGTAACAAATTAATCTTAGTGAATGAATGGA 2460
1981 agattgtcttcttaacgatatctccctcaatatacagcgaacataacacatacaagat 2040
2461 AGATTTGTTTTTACGATATTAATCTCTCCCTTAATTAATGATGATACACATCAAGAT 2500
2041 gagaatagagcaagttcatctcaaaagggttatatgttccagaatccgttctataat 2100
2521 GAGATATATGAGCAAGTTCAATTCAAAAGGTTATATGTTCCAGAAATCCCTTCAATATTA 2580
2101 ctccatggaaccttcaaaaagtgtagaatgaagaatgaatgtagagggttttatcaagaa 2160

Db	1381	TATGAAAAAATGGCAAAAGATTAATAACAGCACTATCAACACTGGAGCGATTCTTATCTGAA	1440
Qy	961	gaagaagagagactttaaanaagctgcgagatccctattgagccaaagaagaatgacata	1020
Db	1441	GAAGCAAGAGAGACTTTTAAAAAGCTGCGAGATTCTATTGAGCCAAAGAAAGAAATACATA	1500
Qy	1021	atcatcttcatctccaagaagaanaagagctctcaaaagaatacaattgatgtagt	1080
Db	1501	ATTCATTCTTTATCTCAAGAAAGAAAAGAGCTTCTAAAAAGAAATCAATTGATGTAGT	1560
Qy	1081	gattcttcatctactgcgaggaagaagctttaaanaagctcaaatgatattcgtgat	1140
Db	1561	GATTTTATCTACTGTGAGGAAAAAGATTTTTAAAAAACCTCAAAATTGATATTCTGTAT	1620
Qy	1141	tcttcatctbaagaagaanaagagctttaaataagatacagttgatatagtagtaactc	1200
Db	1621	TCTTTATCTGAGAAAGAAAAGAGCTTTTAAATAGAAATCAGGTGAGTAGTAGTATCT	1660
Qy	1201	ttaetcgaaanaaganaaagagctttaaanaagctgaaactgtaattcaaccatagat	1260
Db	1681	TTATCTGAAAAAGAAAAAGATTTTTAAAAAGCTGAACCTGATATTCACACATATGAT	1740
Qy	1281	attataaagaattgccaagaatacagaaggtaattgatagtcgcgtcaattaatcttgat	1320
Db	1741	ATTATATCAAAAGTTGCGAAGATACAGAGGGGTTATATGATAGTCCGTCATTAATCTTGAT	1800
Qy	1321	gtaaagaagaagataaanaaggaattccaanaatattgatgcttatacatcatcat	1380
Db	1801	GTAAGAAGAAGAGTATTAAGAGGATATTCAAAATATGATGCTTTATTCATCAATCCAT	1860
Qy	1381	ggaagtaacctgtgacaaataaatttttgtatgaaataatgataatcaatacacttaca	1440
Db	1861	GGAAGTACCTTGATCAATTAATAATTTATTTGTAGAAAAATTAATAATCAATTAACCTTACA	1920
Qy	1441	gcaaccctaggtcggaattagttgatttccaactgataactcaaattaagagtagt	1500
Db	1921	GCAACCTCAGGTGGGATTTAGTTGATTCCACATGATTAATCAATTAATAGAGATT	1960
Qy	1501	ttcaagaattcaaaaaaattccaatataglatcttctagtaactatagattgtgat	1560
Db	1981	TTCAATCAATTCAAAAAAATTTCAAAATATAGATTTTCATGTAAGCTATATATGATGTTGAT	2040
Qy	1561	ataaabaagaagcctgcgataagataaagagcgtttgaaatggaagatccaattatcacca	1620
Db	2041	ATAATATGAAGGCTGCATTAGATTAATGAGCGTTTGAAATGAGATCCAAATTATACACA	2100
Qy	1621	gatactcgacagagataattagaanaatgggaagccttatactcaagaagaacatcgctgc	1680
Db	2101	GATCTCTGACAGAGATTATTAGAAATGGAAGCTTATATTCAAAAGAACATCGGTCGTG	2160
Qy	1681	gaataaagaagattacaanaatlaagcaatccgaaanaagaatatataagattgatgcg	1740
Db	2161	GAATTAAGAAGATGTACAAATAATTAATAAGCAATCGAAAAAGAAATATATGAAGTTGATCGG	2220
Qy	1741	aaagttagtgcacaagaagtaaaatagaatacaaaattcaagaagacagttaaataaat	1800
Db	2221	AAATAGTAGCCCAAGAAGATAAATATAGATACAAAAATTCAAGAAAGCACAGTTAAATTAAT	2280
Qy	1801	caggaatcggaataaagcatlaaggtttaccaaaatatatacaagcctattacattcaagctg	1860
Db	2281	CAGGAATGGAATTAAGCATTAAGGTTTACCAAAATTAACAAACCTTATTAATTCACACCTG	2340
Qy	1861	cataatgatcatgcatccaatctatgagaagtgctatttaatatgaaatgaaaggaa	1920
Db	2341	CATTAATGATATGATCATTCATATTGTAGAAAGTCTTATTTATATATGAATGAATGGAA	2400
Qy	1921	aataatactcaaatgatacttaaaaaaggtacaacaattacttggtcgatggttaatga	1980
Db	2401	AATTAATATTCMAAAGTGATCTTATTAATAAAAGGTATACAAATTAATTGTTGATGATGATGCA	2460
Qy	1981	agattggtctttaccgatatctactccctaataatagctgtaacaatatatacatcaaat	2040

Db	2461	AGATTGTTTTTACCGGATTAATTA	CTCCCTTAATATAGTGGAACAAATATACATCAAGT	2520
Qy	2041	gagataatagcaagttcattcca	aaagggtatagtctcgaatcccggtctatalt	21000
Db	2521	GAGATATATGAGCAAGTCTTC	CAAAAAGGTTATATGTTCCAGAAATCCGTTCTATATTA	25800
Qy	2101	ctccatgaccctcaaaaggtaga	ataaagaaatgaatgaagtggtttatcacaa	21600
Db	2581	CTCCATGACCTTCAAAAGSTG	TAGAAATTAAGCAATGATAGGAGGTTTATACAGAA	26400
Qy	2161	tttgagaaatgctggaatgatta	tgcctgatacttaataagaacaaatcgaatt	22200
Db	2641	TTTGACATGCTGTGATGATTA	TGCTGGATATCTATTAAGATAAGAACCAATCTGATTTA	27000
Qy	2221	gttaacaaatctctaaaaaatc	atcatgatalttttaaggaagaaggagtaatttaactcg	22800
Db	2701	GTTACAAATTTCTAAAAAATTC	ATGATTTATTTTAAGGAAGGAGTAATTTAACTTCG	27600
Qy	2281	tatggagaaacaatgaagcgaa	atcttttcagagaagcccttaagttatgcatctacg	23400
Db	2761	TATGGGAGCAACAATGATAGC	GGAATTTTTTGCAAGAACCTTATAGTTAATGCAATTTACG	28200
Qy	2341	gaacatgctgaacgtttaaaag	ttcaaaaaatgctcgaagaactttccaatttaac	24000
Db	2821	GACCATGCTGAAGTTTAAAG	ATGTTCAAAAAAATGCTCCGAANAACCTTCAATTTATTAC	28800
Qy	2401	gatacaatgaatgcatctatta	actcata	2430
Db	2881	GATCAGATTATAGTCAATTAT	TAACTCATATA	2910

```

RESULT 3
PCT-US94-01624-1
: Sequence 1, Application PC/TUS9401624
:
: GENERAL INFORMATION:
: APPLICANT: Leppia, Stephen H.
: APPLICANT: Klimpel, Kurt R.
: APPLICANT: Ariora, Naveen
: APPLICANT: Singh, Yogendra
: APPLICANT: Nicholas, Peter J.
: TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
: TITLE OF INVENTION: RELATED METHODS
: NUMBER OF SEQUENCES: 31
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
: STREET: Stewart Street Tower, 20th Floor, One Market
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94105
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/01624
: FILING DATE: June 25, 1993
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Weber, Kenneth A.
: REGISTRATION NUMBER: 31,677
: REFERENCE/DOCKET NUMBER: 15280-115
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 543-9600
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3291 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear

```

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus anthracis
FEATURE:
NAME/KEY: CDS
LOCATION: 580..2907
OTHER INFORMATION: /product= "lethal factor"
PCT-US94-01624-1

Query Match 100.0%; Score 2430; DB 5; Length 3291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgagataataaagaattatataaagtaattgtatgtcattttagtaacgaatt 60
DB 481 ATGAAATATMAAAGAAATTTATMAAGTAAATGATGTCATGTTTAAACAGCAATT 540
QY 61 actttagtggtccgcgtcttatcccccgttacaggggcggttcattggtgatgta 120
DB 541 ACTTTGAGTGGTCCCGCTTTATCCCTTGACAGGGGGCGGCGTCAATGATGATGTA 600
QY 121 ggtatgcagctaaagaagaataaataaagatgagaataaagaagaagaagaa 180
DB 601 GGTATGCACGCTMAAAGAGAAAGACAAATTAAGATGAGATAAGACAAAGATGAGAA 660
QY 181 cgaataataacacaggaagagcatttaagaataatcgtaaacacattgtataaatagaa 240
DB 661 CGAAATMAAACACAGAGAGCATTTAAAGCAATTCATGAACCATTTGMAAAATAGAA 720
QY 241 gtaaaaggaggaggaagcgtttaaaaaaggcagcagaagaagcacttggagaagtaacca 300
DB 721 GTAAGAGGAGAGAGCTGTTAAAAAGAGCAGACAAAAAGCTTGACAAAGTACCA 780
QY 301 tctgtagtctttagagatgataaagcaattggagaagaatataattgttgatgtgat 360
DB 781 TCTGATGTTTATGAGATGTATMAAGCAATTGAGAGAAAGATATATTTGTGATGATGAT 840
QY 361 attacaaacatatatctttagaagcattatctgaagaataagaataaataaagaacatt 420
DB 841 ATTAACAAGCATATCTTTTGAAGCATTTCTGAAGATTAAGAAAAAATAAGAAATTT 900
QY 421 tatgggaagaatcgttattacatgaacattatgtatgtaaaagaagaatagaaccc 480
DB 901 TATGGGAAGATGCTTTATTTACATGAACATTTATGATCAAAAAGATGATGAACCC 960
QY 481 gtaactgttaacaaatcttcggaagattatgtagaataactgaaagagcactgaacgt 540
DB 961 GTAATTGTATCCAAATCTTCGGAAGATTATGTGAATAATCTGAAAGGCACTGAACGTT 1020
QY 541 tattatgaataatgtaagaataatcaaggagataatttaagtaaaatlaacaaacat 600
DB 1021 TATTATGAATAATAGTAAAGATTTATCAAGGATTTTAAAGTAAATTAATCAACCATAT 1080
QY 601 cagaatacttttagatgtrattaataacataaataatgcattgatttaagaatgtaagaat 660
DB 1081 CAGAAATTTTAAATGATGATTAATACCATTTAAATGTCATCTGATTCAGATGCAAGAT 1140
QY 661 ctttatttactaatcagcttaaggaacatccacagacttctgtagaattcttgaa 720
DB 1141 CTTTATTTACTATATGAGCTTAAGGAACATCCACACACTTTTCTGTAGAAATTTCTTGAA 1200
QY 721 caaaaatagcaatgaggtacagaagaatttgcgaagccttgcataatlatatcgagca 780
DB 1201 CAAAAATAGCAATGAGGTACAGAAGATTTGCGAAACCTTTGATATTTATTCAGAACCA 1260
QY 781 cagcatcgtatgcttttaagagctttatgacccggaagccttlaatacatcggataaatt 840
DB 1261 CAGCATCGTATGCTTTTACAGCTTTATGACCGGAAAGCTTTTAAATTAATTCAGATGAATTTT 1320
QY 841 aacgacaagaataataatctatccttggagaacttaagaatcaagatgcgtctcaaga 900

DB 1321 AACGAACAGAAATAATCTATCTTGAGAACACTTAAAGATCAACGATCTGTCAAGA 1380
QY 901 tatgaaaaatgaggaaagataaacaacgacatacaaacgtagagcgattcttctcgaa 960
DB 1381 TATGAAAAATGGGAAAGATTAACACACATATCAACCTGAGACGATTTCTTTATCTGAA 1440
QY 961 gaaggaagagactcttaaaaaagctcagaattcttcttagtgccaaagaagaatgacata 1020
DB 1441 GAAGGAAGAGACTTTTAAAAAAGCTGACGATTTCTTATGAGCCAAAGAAAGATGACATA 1500
QY 1021 attcattcttatctcaagaagaagaagcttcttaaaaaagctacaataatgtagtaagt 1080
DB 1501 ATTCAATCTTATATCTCAAGAAAGAAAGAGCTTCTTAAAAACATTCAAATGATAGTAGT 1560
QY 1081 gatttcttactactgaggaagaagcttcttaaaaaagctacaataatgtagatcgtgat 1140
DB 1561 GATTTTATCTACTGAGAAAGAGCTTTTAAAAAGCTACAAATTTGATATTCGTGAT 1620
QY 1141 tcttattctgaagaagaagaagcctttaaataagatacaggtgtagtaagtaacct 1200
DB 1621 TCTTATCTGAGAGAGAAAGAGCTTTTAAATGAAATACAGGTGATGATGATATCT 1680
QY 1201 ttatctgaaaaagaagaagcttcttaaaaaagctgaactgtatctcaaccaatgatc 1260
DB 1681 TTATCTGAAAAAGAAAGAGCTTTTAAAAAGCTGAAACTGTGATTCACACCATATGAT 1740
QY 1261 attaaatcaaaaggttgcagaatgacagaagggttaattgtatgctcgcataatlaactgat 1320
DB 1741 ATTATATCAAGGTTGCAAGATACAGGAGGTTAATGATGATCCGTCGAATTAATCTTGAT 1800
QY 1321 gtaagaagaagcagtaataaaggatattcaaaatattatcttattacttaactcaatcatt 1380
DB 1801 GTTAAGAAAGCATATTAAGAGGATTTCAAAATTTATGCTTTTATTCATCAATTCATT 1860
QY 1381 ggaagtaacctgtgtaaaataattatcttlatgaaataatgaaatlaacaaatlaactca 1440
DB 1861 GGAAGTACCTGTGATCAATTAATTTATTTGATGATGATGATGATGATGATGATGATGAT 1920
QY 1441 gcaacctgaagctgcgagatttagttagtccacgtataatctaaatlaagaagattat 1500
DB 1921 GCAACCTGAGTGGGATTTAGTATTCACATGATTAATCTAAATTAATTAAGAGTAT 1980
QY 1501 ttcaatgaattcaaaaaaatttcaaatatagatatctctgttaactatagattgtgat 1560
DB 1981 TTCAATGAATTTCAAAAAAATTTCAAAATTAATTAATTAATTAATTAATTAATTAATTAAT 2040
QY 1561 ataaatgaagagcctgcattagataatgagcgtttgaaatggaagaatccaattatcca 1620
DB 2041 ATTAATGAAGAGCCTGATTTAGATTAATGATGATGATGATGATGATGATGATGATGAT 2100
QY 1621 gatactgagcagagataattagaanaatggaagcttattatcaaaagaacatcgtgctg 1680
DB 2101 GATACCTGACAGAGATATTTAGAAATGGAAGCTTATTTACAAAGAACATTCGCTG 2160
QY 1681 gaataaagaatgtacaaataatlaagcaatccgaaagaataatataagaatgtgatg 1740
DB 2161 GAATTAAGAGATGTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2220
QY 1741 aaagttagtgcgaagaatlaaataagatacaaaaaatccaagaagaagcttaaatataat 1800
DB 2221 AAGTATGTCGCAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2280
QY 1801 caggaatggaataaagcattaggttaccaaatatacaaaagcttattatcatcaacgtg 1860
DB 2281 CAGGAATGGAATTAAGCATTTAGGTTTACCAAAATTAATTAATTAATTAATTAATTAATTAAT 2340
QY 1861 cataatagatatgcatcccaatattgtagaagaagcttattatataatgaaatgaaatg 1920
DB 2341 CATTAATGATATCATCAATTAATTTAGAAAGCTTATTTAATTAATTAATTAATTAATTAAT 2400
QY 1921 aatataatcaaggtgattctataaaaaaggtaaacaaattacttagttagttagttagt 1980

Db	2401	AATPATTTTCAAGTAGTCTTATTAATAAAAGTAACAAATTAAGTATAGTATGTAATGA	2460
Qy	1981	agatttggttttcgcgatatctctccctaataagctggaacaataaacaatcaagat	2040
Db	2461	AGATTGTGTTTTACCAATATTAATCTCCCTTAATATAGCTGAACAATATACATCAAGAT	2520
Qy	2041	gaagatalatgagaagttcatctcaaaagggtatatatgttccagaatcccgctctata	2100
Db	2521	GAGATATATGAGCAAGTTCATTCCAAAGGGTTATATGTTCCAGAATCCGTTTATATTA	2580
Qy	2101	ctccatggaacctccaagggtgtagaataaaggaatgatagtgaggtttatacaaga	2160
Db	2581	CTCCATGAGACCTTCAAAAGGTGATGAATTAAGAAATGATAGAGAGGTTTATACAGAA	2640
Qy	2161	tttggaacgtctgtgatatgtctgtgatactcatagataaagaaccaatcggattta	2220
Db	2641	TTTGGACATCTGTGATGATTTATGCTGTGATATCTATTAAGATMAAACCAATCTGATTTA	2700
Qy	2221	gttacaacatctcaaaaatcaltcgtatalttttaaggaagaaaggagtaattcaactcg	2280
Db	2701	GTTACAAATTTCTAAAAAATTCATTCGATATTTTAAAGAAAGGAGGATTAATTAACCTCG	2760
Qy	2281	tatggaggaacaaatgaacggaatttttggagaagccttagtgtaatacatctcag	2340
Db	2761	TATGGAGAACCAATATGAAGGGGAATTTTGTGAGAAAGCCTTAGGTATATCATTCTACG	2820
Qy	2341	gaccatgctgaacgttttaaaagtccaanaaaatgtctcgaaaacatctcaattatctaac	2400
Db	2821	GACCATGCTGAAGCTTTAAAGATTCACAAAATAATGCTCCGAAACATCTTCATTTATTAAC	2880
Qy	2401	gatcagaatgaagctatcatcttaactcatca	2430
Db	2881	GATCAGATTAGTTCATTATTAACCTCATTA	2910

```

?      LENGTH: 1368 base pairs
?      TYPE: NUCLEIC ACID
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: DNA (genomic)
?      FEATURE:
?      NAME/KEY: CDS
?      LOCATION: 1..1368
?      OS-08-021-601-5

```



```

1 APPLICANT: Leppla, Stephen H.
2 APPLICANT: Klimpel, Kurt R.
3 APPLICANT: Atoria, Naveen
4 APPLICANT: Singh, Yogendra
5 APPLICANT: Nichols, Peter J.
6 TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
7 TITLE OF INVENTION: Related Methods
8 NUMBER OF SEQUENCES: 35
9 CORRESPONDENCE ADDRESS:
10 ADDRESSEE: Townsend and Townsend and Crew LLP
11 STREET: Two Embarcadero Center, Eighth Floor
12 CITY: San Francisco
13 STATE: California
14 COUNTRY: USA
15 ZIP: 94111-3834
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Floppy disk
18 COMPUTER: IBM PC compatible
19 OPERATING SYSTEM: PC-DOS/MS-DOS
20 SOFTWARE: PatentIn Release #1.0, Version #1.30
21 CURRENT APPLICATION DATA:
22 APPLICATION NUMBER: US/08/082,849B
23 FILING DATE: 25-JUN-1993
24 CLASSIFICATION: 514
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: US 08/021,601
27 FILING DATE: 12-FEB-1993
28 ATTORNEY/AGENT INFORMATION:
29 NAME: Weber, Kenneth A.
30 REGISTRATION NUMBER: 31,677
31 REFERENCE/DOCKET NUMBER: 15280-161-1
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE: (415) 576-0200
34 TELEFAX: (415) 576-0300
35 INFORMATION FOR SEQ ID NO: 5:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 1368 base pairs
38 TYPE: nucleic acid
39 STRANDEDNESS: single
40 TOPOLOGY: linear
41 MOLECULE TYPE: DNA (genomic)
42 HYPOTHEetical: NO
43 ANTI-SENSE: NO
44 ORIGINAL SOURCE:
45 ORGANISM: Bacillus anthracis
46 FEATURE:
47 NAME/KEY: CDS
48 LOCATION: 1..1368
49 OTHER INFORMATION: /product= "Lf(1-254)--Tr--PE(401-602)"
50 US-08-082-849B-5

```

Db	241	ATATATATTGTGGATGGTGATATTACAAACATATATCTTTAGAGCATTTATCTGAAGAT	300
QY	400	aagaaataataaagacatltatlggaaagatgcttatlacatgacaatlgtalat	459
Db	301	AAGAAAAAATAAAGACATTTATGGGAAGATGCTTTATTACTGAGCACTTNTGTAT	360
QY	460	gcaaaagaagataltgaaacccgactctgttaatccaatcttcggaagatatgtagaac	519
Db	361	GCAAAAGACGATATGAAACCCGCTCTGTATCCAACTTCGGMAGATATGTAGAAAT	420
QY	520	actgaaagagcagcgaacggttatctatgaaataagtaagataltatcaagaagatltta	579
Db	421	ACTGAAAGGACACTGAGACGTTTATTATGMAANTAGGTAGATATTATTAAGGATATTTTA	480
QY	580	agtaaaatlaacacatcatcagaaatlcttagatgtatcaatataacatlaaataatgca	639
Db	481	AGTAAATTAATTAACAACCATACAGAAATTTTATAGATGTATTAATACCATTAATAATGCA	540
QY	640	tctatltcagatlggacaaagatcttatlactaatacagcttaagaaatccacagac	699
Db	541	TCTGATTCAGATGAGCAAGATCTTTTATTACTAATAGCTTAAGGAATCCACAGAC	600
QY	700	ttctctgtagatctcttggaaacaaatagcaatgaggtacaagaagratlttgaaagct	759
Db	601	TTTTCTGTAGATTTCTTGGAACAAAATAGCATGAGGTACAGAGATTTGGGAAAGCT	660
QY	760	tttgcataltcatcagacccacagacatcgtgatgttlttaagcttltgacccggaagct	819
Db	661	TTTGCAATTTATATCGAGCCACACATCGTCAATGTTTACAGCTTTATGACCGGAAAGCT	720
QY	820	tltaattacatlgatlaaatttaacgacaagaataaatactatccttggaaag	871
Db	721	TTTATATTACATGATTAATTAATTAACGAAAGAAATTAATCTACTCGCGCAGC	772
RESULT	6		
	PCT-US94-01624-5		
	Sequence 5, Application PC/TUS9401624		
	GENERAL INFORMATION:		
	APPLICANT: Leppla, Stephen H.		
	APPLICANT: Klimpel, Kurt R.		
	APPLICANT: Arota, Naveen		
	APPLICANT: Singh, Yogendra		
	APPLICANT: Nichols, Peter J.		
	TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND		
	NUMBER OF INVENTION: RELATED METHODS		
	NUMBER OF SEQUENCES: 31		
	CORRESPONDENCE ADDRESS:		
	ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE and CREW		
	STREET: Stewart Street Tower, 20th Floor, One Market		
	STREET: Plaza		
	CITY: San Francisco		
	STATE: CA		
	COUNTRY: USA		
	ZIP: 94105		
	COMPUTER READABLE FORM:		
	MEDIUM TYPE: Floppy disk		
	COMPUTER: IBM PC compatible		
	OPERATING SYSTEM: PC-DOS/MS-DOS		
	SOFTWARE: PatentIn Release #1.0, Version #1.25		
	CURRENT APPLICATION DATA:		
	APPLICATION NUMBER: PCT/US94/01624		
	FILING DATE: June 25, 1993		
	CLASSIFICATION:		
	ATTORNEY/AGENT INFORMATION:		
	NAME: Weber, Kenneth A.		
	REGISTRATION NUMBER: 31,677		
	REFERENCE/DOCKET NUMBER: 15280-115		
	TELECOMMUNICATION INFORMATION:		
	TELEPHONE: (415) 543-5040		
	TELEFAX: (415) 543-5043		
	INFORMATION FOR SEQ ID NO: 5:		

[illegible]

```

RESULT 7
US-08-021-601-7
: Sequence 7, Application US/08021601
: Patent No. 5591631
: GENERAL INFORMATION:
: APPLICANT: Leppla, Stephen H.
: APPLICANT: Klimpel, Kurt R.
: APPLICANT: Nichols, Peter J.
: APPLICANT: Arora, Naveen
: APPLICANT: Singh, Yogendra
: TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
: TITLE OF INVENTION: RELATED METHODS
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Needle & Rosenberg, P.C.
: STREET: 133 Carnegie Way, Suite 400
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: USA
: ZIP: 30303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/021,601
: FILING DATE: 19930212
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Spratt, Gwendolyn D.
: REGISTRATION NUMBER: 36,016
: REFERENCE/DOCKET NUMBER: 1414, 057
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 404/688-0770
: TELEFAX: 404/688-9880
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1425 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1416
: US-08-021-601-7

Query Match 31.4%; Score 762.4; DB 1; Length 1425;
Best Local Similarity 99.9%; Pred. No. 2.2e-146;
Matches 763; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 100 gcggcgagtcacatgagtgatgaagtatgcacgtaaagaagagaagaaaaataaagatgag 159
Db 10 GCGGGCGGTCACATGATGATGTAGTATGCACTAAAGAGAAAGAAAAATAATGAATGAG 69

QY 160 aataagagaaaagatgtaaacgaacgaataaacaacaggaagcatttaaggaaatcatg 219
Db 70 AATAAGAGAAAAGATGAAGAAAGCAAGAAATAAACACAGGAAGACCAATTAAAGCAATTCATG 129

QY 220 aaacgacttgtaaaaaatgaagaataaaggggaggaagctgttaaaaaaaggacagagaa 279
Db 130 AAACGACTTTGTAATAATGAAAGTAATAAGGGGAGAAAGCTGTTAATAAAGAGCGACAGAA 189

QY 280 aagcctacttgaaagatccacatctgctgttcttagagatgtataaagcaattggagaaag 339
Db 190 AAGCCTACTTGAAAGATGCCATCTGATGTTTTAAAGATGTATAAAGCAATTGAGAGAAAG 249

QY 340 atatatatttgatgtgtgatatatcaaaaatatatccttgaagacattatctgaagat 399
|||||

```

Db	250	ATATATATGTGGATGGTGATATTACAAACACATATATCTTTAGACGATTATCTGAAGAT	369
Qy	400	aagaaaaataaagaacattttagygaagaagtccttattacaatgaacataltgatat	459
Db	310	AAGAAAAAATAAAGACATTTTATGGGAAAGATGCTTTATTACATGAACTTATGTATAT	369
Qy	460	gcaaaagaagataltgaacccgactctgtaatccaatcttcggaagttatgtagaaat	519
Db	370	GCAAAAGAGATATGAACCCGTACTTGTAATCCAACTTCGGAAAGATTATGTAGAAAT	429
Qy	520	actgaaaagcagcaacgcttactatcgaataagylaagatatattacaaggatattta	579
Db	430	ACTGAAAAGCAGCAACGCTTTATTATGAATAAGTAAAGATATTATCAAGGATATTTTA	489
Qy	580	agtaaaatlaatcaaccatalcagaaaattttagatglatlaaataccatlaaanaatga	639
Db	490	AGTAAATTTATCAACCATATCAGAATTTTATAGATGTATTAATACATTAAATAATGCA	549
Qy	640	ctgatccagttgacaagaatccttatttactcaatcagctttagaacaatcccaagac	699
Db	550	TCTGATTCAGATGGAACAAGATCTTTTATTACTAATACGTTTAAAGCAATCCACAGAC	609
Qy	700	ttctctgtaagaatctcttggaacaaatagcaatgagtgacaagaagratcttggaagct	759
Db	610	TTTTCTGTGAGATTCCTTGGAACAAATAAGCAATGAGGTAAAGAAAGTATTGGGAAAGCT	669
Qy	760	tttgcatattatalcgagccacagcatcgtgaltgttttacaagctttagcaccggaagct	819
Db	670	TTTGCAATATTTATTCGAGCCACAGCATCGTGATGTTTATACAGCTTTATGACCGGAAAGCT	729
Qy	820	tttaattacatggttaaatlaaagaacaaagaataataatctatc	863
Db	730	TTTAATTACATGATGAATTAATTTAAAGCAACGAATAATATCTAAC	773
RESULT 8			
US-08-082-849B-7			
Sequence 7, Application US/08082849B			
Patent No. 5677274			
GENERAL INFORMATION:			
APPLICANT: Leppla, Stephen H.			
APPLICANT: Kilmpel, Kurt R.			
APPLICANT: Alore, Naveen			
APPLICANT: Singh, Yogendra			
APPLICANT: Nichols, Peter J.			
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and			
NUMBER OF SEQUENCES: 35			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Townsend and Townsend and Crew LLP			
STREET: Two Embarcadero Center, Eighth Floor			
CITY: San Francisco			
STATE: California			
COUNTRY: USA			
ZIP: 94111-3834			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: Patentin Release #1.0, Version #1.30			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/082,849B			
FILING DATE: 25-JUN-1993			
CLASSIFICATION: 514			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 08/021,601			
FILING DATE: 12-FEB-1993			
ATTORNEY/AGENT INFORMATION:			
NAME: Weber, Kenneth A.			
REGISTRATION NUMBER: 31,677			
REFERENCE/DOCKET NUMBER: 15280-161-1			
TELECOMMUNICATION INFORMATION:			
TELEPHONE: (415) 576-0200			

:		TELEFAX:	(415) 576-0300	
:		INFORMATION FOR SEQ ID NO:	7:	
:		SEQUENCE CHARACTERISTICS:		
:		LENGTH:	1425 base pairs	
:		TYPE:	nucleic acid	
:		STRANDEDNESS:	single	
:		TOPOLOGY:	linear	
:		MOLECULE TYPE:	DNA (genomic)	
:		HYPOTHETICAL:	NO	
:		ORIGINAL SOURCE:		
:		ORGANISM:	Bacillus anthracis	
:		FEATURE:		
:		NAME/KEY:	CDS	
:		LOCATION:	1..1416	
:		OTHER INFORMATION:	/product= "LF(1-254)--TR--PE(398-613)"	
:		US-08-082-849B-7		
: Query Match			31.4%;	Score 762.4; DB 1; Length 1425;
: Best Local Similarity			99.9%;	Pred. No. 2.2e-146;
: Matches 763; Conservative			0; Mismatches 1; Indels 0; Gaps 0;	
OY	100	gcggcgccgctcatcgtgcgatgtagtattgccacgtaaaagagaagaataaatgaatgag	159	
Db	10	GGGCGCGCTCATGTGCATGTAGTATGCCACGTAAAGAACAATAAATTAAACATGAG	69	
OY	160	aataagagaaaaagatcgaagaaagcaatataaaacacaaggagaagcatttaaaggaatactg	219	
Db	70	AATAAGAGAAAAAGATCAGAACGAATATAAACACAGGAAGACATTTAAGGAATCATG	129	
OY	220	aaacacattgtiaaaaatagaaglaaaaggggaggaagctgtttaaaaaaaggcacagaa	279	
Db	130	AAACCAATTGTAAAANTAGAACTAAAGGGGAGAAAGCTTTAAAAAAGAGCGCACAGAA	189	
OY	280	aagcactcttgagaagtcccaccttatgttttagagatgtaataaagccaattggaagaaag	339	
Db	130	AAGCACTTGAGAAAAAGTACCATCTGATGTTTTAGAGATGATTAAGAACAATTGGAGAAAG	249	
OY	340	atataatcttgtatcgtgfatatatacaaacaatatatctttagaagcattacctaagat	399	
Db	250	ATATATATTGTTCGATGCGTGATATTACAAACAATATCTTTAGMAGCATATTACTGAAGT	309	
OY	400	aagaaanaaaaataaaagacattatvgggaagaatgcttattacacatgaaacattatgatat	459	
Db	310	AAGAAAAAAAAATPAAAAGACATTTATGGGAAGAATGCTTTATTACATGAACATTTATAT	369	
OY	460	gcaaaaagaagatatgaaccocglacttgttaaccaatcttcggaagattatglaaaaaat	519	
Db	370	GCAAAAAGAGGTATGAAACCCGTACTTGTAAACCAATCTCCGAAGATTATCTPAAGAAAT	429	
OY	520	actgnaaaggcactgaacgctttattcatgtaaaataggtaagataatacaaggagatttta	579	
Db	430	ACTGAAAAAGCCACTAAACGTTTATTATATGAATATAGTAAGATATTITCAAAGGATRTTTTA	489	
OY	580	aglaaaatlaalcacaacatalcogaanaatlitraga tgfattaalaatcacattiaaaaalgc	639	
Db	490	AGTAATAATTATATCAACCAATACAGAAATTTTAGATATGATTAATAATACATTAAAAATGCA	549	
OY	640	tctgaattcagaatggaacaagatcctttatttactaacacagcttaaggaaacatccccagac	699	
Db	550	TCTGATTAGATGCGCAAAAGATCTTTTATTACTAATACGCTTAAGCAACATCCCCAGAGC	609	
OY	700	ttttcttgtaagaattcttgaaacaanaataggcaatlgatgafacaagaagattttggcgaagct	759	
Db	610	TTTTCTGTAGAAATTTCTTGAAACMAAATATGCAATGAGGTACAAAGAAATTTTGGCAAGCT	669	
OY	760	tttgcatattatcatcagaccacagacatcgtlgaigtctttacagccttacgaccggaagct	819	
Db	670	TTTGATATATTATPATCGACCAACAGCATGCTGATGTTTTAACAGCTTTATGCAACCGGAAGCT	729	
OY	820	tttaatcatacgtataaatattaaegaaacaagaataaatactatc 863		
Db	730	TTTTATTTACATGATTAATTTTACCACMACAAAGAAATTAATTTATCAC 773		

TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1524 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1524
US-08-021-601-9

Query Match 31.4%; Score 762.4; DB 1; Length 1524;
Best Local Similarity 99.9%; Pred. No. 2.2e-146;
Matches 763; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
OY 100 gcggcggtcatggtgtagtgcacgttaaaagagaagaataaagatgag 159
    |||||||
DB 1 gcggcggtcatggtgtagtgcacgttaaaagagaagaataaagatgag 60
    |||||||
OY 160 aataagagaaaagatgaagacgaataaaacacaggaagcatttaaggaatcatg 219
    |||||||
DB 61 AATAAGAGAAAAGATGAGAACGAATATAAACACAGGAAGACATTATAAGGAATCATG 120
    |||||||
OY 220 aaacacattgttaaaaatagaagtaaaagggggaagctgttaaaaagaagcagcagaa 279
    |||||||
DB 121 AAACACATTGTATAAATAGAAATGAAGTAAAGGGAGAAAGCTGTAAAAAAGAGCAGAGAA 180
    |||||||
OY 280 aagctacttgagaagaatgccatcgtatgttttaagagatgataaagcaattggagaag 339
    |||||||
DB 161 AAGGACTTGAGAAAAGTACCATCTGATGTTTAGAGATGTATTAAGCAATTTGAGAGAAAG 240
    |||||||
OY 340 atataattgtgagatgagatataaacaaacatacatcttgaagcattatcctgaagat 399
    |||||||
DB 241 ATATATATTGTGATGATGATGATTAACAACATATATCTTTGAAGCATATCTGAAGAT 300
    |||||||
OY 400 aagaaaataaataaagaacattatgggaagatgcttataacatgaacattatgatat 459
    |||||||
DB 301 AAGAAAAAATTAAGACATTTATGGAAGATGCTTTATATACATGACATTTATGATATAT 360
    |||||||
OY 460 gcaaaagagaatgaagaccgctactgttaacccaatcttcggaagattatagaagat 519
    |||||||
DB 361 GCAAAAGAAAGATATGAACCCGCTACTTGTATTCATCTTCGGAAGATTATGAGAAAT 420
    |||||||
OY 520 actgaaagagcactgaagcttatataatgaataggaagataatacaaggagatat 579
    |||||||
DB 421 ACTGAAAAGGCCTGACAGTTTATGAAATAGGTAAGATATTTATCAAGGATATTTTA 480
    |||||||
OY 560 agtaaaatcaatcaaccatatacagaatctttagatgtaataaacaataaaatgca 639
    |||||||
DB 481 AGTAAATTAATCAACCATATCAGAAATTTTATGATGTATTAATCAATTAATAATGCA 540
    |||||||
OY 640 tctgattgaatggaagaagatctttattactaatacagcttaaggaataccccaagc 699
    |||||||
DB 541 TCTGATTAGATGAGACAAAGATCTTTTATTACTTAATCAGCTTAAGAACATCCACAGAC 600
    |||||||
OY 700 ttctctgtagaattcttgaacaataagcaatgaggtlaacgaagatttcgaaagct 759
    |||||||
DB 601 TTTTCTGTAGAATTTCTTGSAAACAAAATAGCATGAGTCAAGAAATATTTCGAAAGCT 660
    |||||||
OY 760 ttgcataattatatacgaaccacagcatcgtagtcttcaagcttataagcaccggaagct 819
    |||||||
DB 661 TTTGATATTATATTCAGACACAGCATGCTGATGTTTACACCTTATATCACCAGAACT 720
    |||||||
OY 820 ttaattacatgataaatttaagaaacagaataaataactc 863
    |||||||
DB 721 TTTAATTACATGATTAATTTAAGCAACAGAAATTAATCTAAC 764
```

```
RESULT 11
US-08-082-849B-9
Sequence 9, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Kilmpel, Kurt R.
APPLICANT: Arore, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
FAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1524 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus anthracis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1524
OTHER INFORMATION: /product= "LF(1-254) --TR--PE(362-613)"
US-08-082-849B-9
Query Match 31.4%; Score 762.4; DB 1; Length 1524;
Best Local Similarity 99.9%; Pred. No. 2.2e-146;
Matches 763; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
|||||
Db 181 AAGCTACTTGAGAAAGTACCTGATGCTTTTAAAGATGATTAAGCAATTTGGAGGAAG 240
Qy 340 atacatactgtggaatggtatattacaaacatactcttaagaagcaatccgaagat 399
Db 241 ATATATTTTGGATGGTGGATTTACAAAACATATATCTTTAGAAAGCATTTCTGAAGAT 300
Qy 400 aagaaaaaaataaagaacattatgggaagaatgcttattacatgaacattatgatat 459
Db 301 AAGAAAAAATTAAGACATTTATGGAAGATGCTTTATTAACATTAAGATTAATATAT 360
Qy 460 gcaaaaagaagatggaacccgtctgttaacaaatcttcggaagatgatagtagaat 519
Db 361 GCAAAAGAGATATGAGACCGTACTGTAATCCAATCTTGGAAGATTAATGGAATAAT 420
Qy 520 actgaagaagcagcgaacgtttatatagaataggaatataatcaaggacatttta 579
Db 421 ACTGAAAAGCAGTCAACGCTTTATTAAGAAATAGTAAGATTAATCAAGGATATTTTA 480
Qy 580 agtaaaatcaatcaacacatacgaanaatltagatgaltataatcacatataaatagca 639
Db 481 AGTAAATTAATCAACCATATCAAGAAATTTTATGATGATTAATTAACATTAATAATGCA 540
Qy 640 tctgattcagatggaagaagatctttattactaatacagcttaagaagacatcccaagac 699
Db 541 TCTGATTCAGATGACACAGATCTTTTATTTACTAATCAAGCTTAAGACATCCACAGAC 600
Qy 700 ttctcgtagaatctctggaacaaatagcaatgaaggtacagaagatatttgcgaagct 759
Db 601 TTTTCTGTAGAAATCTTGGAACAAATATAGCAATGAGTACAAAGATTTTGCAGAAAGCT 660
Qy 760 ttgcatatataatcgagccacagacatcgatgatttcttaagcttatacgacggaagct 819
Db 661 TTTTGCAATTAATATACGACCCACAGCATCTGATGCTTTTACAGCTTTATGACCGGAAGCT 720
Qy 820 tttaattacatgataaatttaacgaacagaanaataaataatc 863
Db 721 TTTAATTAATCAATGATTAATTAACGAACAGAAATTAATTAATCAAT 764

RESULT 12
PCT-US94-01624-9
; Sequence 9, Application PC/TUS9401624
; GENERAL INFORMATION:
; APPLICANT: Leppia, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Airoa, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
; STREET: Steuart Street Tower, 20th Floor, One Market
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01624
; FILING DATE: June 25, 1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-115
```

```
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1524 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus anthracis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1524
; OTHER INFORMATION:
; OTHER INFORMATION: "L(1-254)--TR--PE(362-613)"
PCT-US94-01624-9

Query Match 31.4%; Score 762.4; DB 5; Length 1524;
Best Local Similarity 99.9%; Pred. No. 2,2e-146;
Matches 763; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 100 gggggcgatcagtgatgtagtgcagtaaaagagaagaagaataaagatgag 159
Db 1 GCGGGCGGTCAATGCTGATGTTAGTATGACGTAAAGAGAAAGAAATTAAGATGAG 60
Qy 160 aataaggaanaagatggaacacgaanaataaacaagaagaagacattaaagaaatcag 219
Db 61 AATAAGGAANAAGATGAGAAAGCAAAATTAACACGGAAGAGCTTTTAAGGAATATCTG 120
Qy 220 aaacacattgtaaaaatagaagtaaaaggagggaagctgttaaaaaaggcagcagaa 279
Db 121 AAACACATTTGTAATAATAGAAAGTAAAGGGGAGGAGAGCTTTAATAAAAGACGACGAA 180
Qy 280 aagctacttgagaagatcacatcgtatgctttagagatgataaagcaattggaagaag 339
Db 181 AAGCTACTTGAGAAAGTACCATCTGATGCTTTTAAAGATGATTAAGCAATTTGGAAGAA 240
Qy 340 atacatactgtggaatggtatattacaaacatactcttaagaagcattatccgaagat 399
Db 241 ATATATTTTGGATGGTGGATTTACAAAACATATATCTTTAGAAAGCATTTCTGAAGAT 300
Qy 400 aagaaaaaaataaagaacattatgggaagaatgcttattacatgaacattatgatat 459
Db 301 AAGAAAAAATTAAGACATTTATGGAAGATGCTTTATTAACATTAAGATTAATATAT 360
Qy 460 gcaaaaagaagatggaacccgtctgttaacaaatcttcggaagatgatagtagaat 519
Db 361 GCAAAAGAGATATGAGACCGTACTGTAATCCAATCTTGGAAGATTAATGGAATAAT 420
Qy 520 actgaagaagcagcgaacgtttatatagaataggaatataatcaaggacatttta 579
Db 421 ACTGAAAAGCAGTCAACGCTTTATTAAGAAATAGTAAGATTAATCAAGGATATTTTA 480
Qy 580 agtaaaatcaatcaacacatacgaanaatltagatgaltataatcacatataaatagca 639
Db 481 AGTAAATTAATCAACCATATCAAGAAATTTTATGATGATTAATTAACATTAATAATGCA 540
Qy 640 tctgattcagatggaagaagatctttattactaatacagcttaagaagacatcccaagac 699
Db 541 TCTGATTCAGATGACACAGATCTTTTATTTACTAATCAAGCTTAAGACATCCACAGAC 600
Qy 700 ttctcgtagaatctctggaacaaatagcaatgaaggtacagaagatatttgcgaagct 759
Db 601 TTTTCTGTAGAAATCTTGGAACAAATATAGCAATGAGTACAAAGATTTTGCAGAAAGCT 660
Qy 760 ttgcatatataatcgagccacagacatcgatgatttcttaagcttatacgacggaagct 819
Db 661 TTTGCAATTAATATACGACCCACAGCATCTGATGCTTTTACAGCTTTATGACCGGAAGCT 720
```

Oy 820 tttaatacatgataaatttaacgacagaagaataatcatc 863
Db 721 TTTAATTACATGATTAATTTAACGACAGAAATAATCACTAAC 764

RESULT 13
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/232.463
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300. 6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29, 768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PT29pt-F15
US-08-232-463-14

Query Match 3.8%; Score 91.6; DB 1; Length 7218;
Best Local Similarity 2.6%; Pred. No. 3.9e-10;
Matches 10; Conservative 255; Mismatches 119; Indels 0; Gaps 0;
Oy 89 ttgtacaggggcgggcgtcattggtatgtatgcacgtaaaagaagaagaanaa 148
Db 1440 TCGTACRR 1381
Oy 149 ataagaatgagagataagaagaagaagaacgaataaacaacaggaagacattaa 208
Db 1380 RRR 1321
Oy 209 aggaatcatgaacacattgttaaaaatagaagttaaaggaggagagctgttaaaaag 268
Db 1320 RRR 1261
Oy 269 aggcagcagaaaagctactgtgaaagtacacatcgtatgtttaagatgataaagcaa 328
Matches 10; Conservative 255; Mismatches 119; Indels 0; Gaps 0;

Db 1260 RRR 1201
Oy 329 ttggaagaaagatatatgtgtgattcacaacataatcttgaagcat 368
Db 1200 RRR 1141
Oy 389 tatctgaagaataaagaacacattatcgtgaagaatgcttatacagaac 448
Db 1140 RRR 1081
Oy 449 atatgtatatacgaagaagaat 472
Db 1080 RRR 1057

RESULT 14
US-08-973-462-2
; Sequence 2, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DROULHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973.462B
; EARLIER FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5361
; TYPE: DNA
; ORGANISM: P. falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(5361)
US-08-973-462-2

Query Match 2.7%; Score 66.4; DB 4; Length 5361;
Best Local Similarity 42.6%; Pred. No. 4.7e-05;
Matches 468; Conservative 0; Mismatches 621; Indels 9; Gaps 2;
Oy 110 atgtgtatgtatggtatgcacgttaaaagaagaagaataaagaatgaagaaga 169
Db 3902 atgaagttgtatgaattaaagaatgtcgaagaagaacagatcgaaagaatctgtattaa 3961
Oy 170 aagatgaagaacgaataaacaacaggaagagcatttaagaagaatcatgaacacattg 229
Db 3962 aagatcttgaagaagatatattaaagaagtaaaagaacacgaacttgaaagtga 4021
Oy 230 taanaaatagaagtaaaaggaggagcgtttaaaagaaggagcagaagaacatctctg 289
Db 4022 tttaagaagattataaagaattaaacatttgaacagatatcttaagaagaanaaag 4081
Oy 290 agaaaatcacatctgttctttagaagtataaagaacatttgaaggaagatatattg 349
Db 4082 aaatagaanaaagatcatcttgaanaaattcgaagaagaagcttgaagaatacaagatctt 4140
Oy 350 tggatgtgatatcaacaacatatctttagaagcattatctgaagaatgaagaanaa 409
Db 4141 -gaagcagatatataaagaagatcttcatcttagaagttgaagaanaaataag 4198
Oy 410 taaagaacattttggaagaagatgtcttattacatgaacattatgtatgtcgaagaag 469
Db 4199 aagaagttacacgaattaaagaagaggtgaacacataataagtggtgacgcataata 4258
Oy 470 gatattgaaccgtacttgaatcgaatcttgcgaagaatattatgataaatacgaagaag 529
Db 4259 aaggtttggaagaagatgtatttagaagaagtgatgatttaagaagaagatattatgaca 4318

```

OY 530 cactgaagcttattatgatgaataagtgatgatctatcaaggagatatttaagtaaat 589
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4319 tgttaagggaagatcttgatatttagggatattgataaggaaagcttgaagaattaca 4378
OY 530 atcaaccatcagaattatttaggtgttataatcaccttaaaaatgcatctgattcag 649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4379 caaaactctggaaaaagcttgatctccttaaaagatgtttactctgtgcatatgcat 4438
OY 650 atggaacagaactcttatttacttaactcagcttaagaaacatcccaagactttctgtg 709
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4439 atgaagaacaaaatgaaaaaagaaaaaaagccaaagacttaagtgtgaagaatattat 4498
OY 710 aattcttggaacaaaatagcaatggtggtacgaagaagattttcgaaagctttgcatatt 769
Db 4499 taagaagaaggttcaagaagaaccaaagaaaaaatacaaaaaaagaagaagtttg 4558
OY 770 atatcgagcgcaagcatctgtgatgttttcaagctttatgcacgggaagctttaattaca 829
Db 4559 atatttaagatataaggaaaccaaagatgatatgtagaagttgaagtgaagaattgaata 4618
OY 830 tggataaatttaacgaacaaagaataatctctccttggaaagaaacttaagaatcaagg 889
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4619 tagaagaagatgtgaagaagaatacaagaagaatataagaagaataaagtttgaagata 4678
OY 890 tgcgtcgaagatcatgaaaaatggtggaagaataaacaagcactcaacactgagcgatt 949
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4679 t-----agatgaagatataagttgaagatattgtgtgaagacaaagaatgaagttatgatt 4732
OY 950 ctttactgaaagaaggaaggaactttaaaaaagctgcagattcctattgagccaaga 1009
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4733 taattgtccaaaaaagagaacgcacttgaaaagtttaagcgaaaaaagaataattaga 4792
OY 1010 aagatgacataattcatcttcttattctcaagaagaagaagaagcttcttaaaagaataca 1069
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4793 aaaaagttgaagaaggtgtctagttgtcttaaaaaaacgtagacgaagtaatgaatag 4852
OY 1070 ttgatagtagtatttttattcttactctgaggaaaaaaggtttttaaaaaagctacaatg 1129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4853 ttcaaaaaatgataaagagcttgataaagaagtatctaaagctttgatacaaaaaatg 4912
OY 1130 atattcgattcttattctgaaagaagaagaagactttaatataatgaacaggtgata 1189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4913 agttacttaagtcttcaaaaaacaatacaagattcttttagtaagttaaaaacttcgtaa 4972
OY 1190 gtgaatccttactctg 1207
    ||| ||| ||| |||
Db 4973 aaaaatataagatttg 4990

RESULT 15
US-08-973-462-1
: Sequence 1, Application US/08973462B
: Patent No. 6191270
: GENERAL INFORMATION:
: APPLICANT: DRULHE, PIERRE
: APPLICANT: DOUBERSTES, PIERRE
: TITLE OF INVENTION: MALARAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
: FILE REFERENCE: 0660-0125-0 PCT
: CURRENT APPLICATION NUMBER: US/08/973, 462B
: CURRENT FILING DATE: 1998-02-06
: EARLIER APPLICATION NUMBER: PCT/FR96/00894
: EARLIER FILING DATE: 1996-06-12
: EARLIER APPLICATION NUMBER: FR 95/07007
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 6152
: TYPE: DNA
: ORGANISM: P. falciparum
US-08-973-462-1

```

[illegible]

Mon' Dec 3 10:23:17 2001

[illegible]

Search completed: December 2, 2001, 13:51:02
Job time: 117 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 2, 2001, 13:49:05 ; Search time 129.57 Seconds
(without alignments)
16078.570 Million cell updates/sec

Title: US-09-747-521-1
Perfect score: 2430
Sequence: 1 atgataataaaaaagaatt.....agtcattatataactacataa 2430

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N.Geneseq_1101:*
2: /SID2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
3: /SID2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
4: /SID2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
5: /SID2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
6: /SID2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
7: /SID2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
8: /SID2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
9: /SID2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
10: /SID2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
11: /SID2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
12: /SID2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
13: /SID2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
14: /SID2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
15: /SID2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
16: /SID2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
17: /SID2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
18: /SID2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
19: /SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
20: /SID2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
21: /SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
22: /SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
23: /SID2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2430	100.0	2430	22	AAC86015	Wild type B. anthr
2	2430	100.0	3294	15	AAO70179	Sequence encoding
3	767.6	31.6	1455	22	AA500250	Lfn-Bcl-Xl apolpos
4	762.4	31.4	1368	15	AAO70181	Lfn(1-254)-TR-PR
5	762.4	31.4	1425	15	AAO70182	Lfn(1-254)-TR-PR
6	762.4	31.4	1524	15	AAO70183	Lfn(1-254)-TR-PR
7	211.2	8.7	936	22	AA58252	Oligonucleotide D1
8	211.2	8.7	936	22	AA58254	Oligonucleotide D1
9	211.2	8.7	936	22	AA58257	Oligonucleotide D1
10	211.2	8.7	936	22	AA58259	Oligonucleotide D1
11	211.2	8.7	936	22	AA58262	Oligonucleotide D2

C	12	211.2	8.7	938	22	AA58255	Oligonucleotide D1
	13	209.4	8.6	936	22	AA58252	Oligonucleotide D1
	14	209.4	8.6	936	22	AA58254	Oligonucleotide D1
	15	209.4	8.6	936	22	AA58257	Oligonucleotide D1
	16	209.4	8.6	936	22	AA58259	Oligonucleotide D2
	17	209.4	8.6	936	22	AA58262	Oligonucleotide D2
	18	209.4	8.6	938	22	AA58255	Oligonucleotide D1
	19	157.8	6.5	2990	11	AAO04123	Adenyl cyclase gen
	20	98.4	4.0	1686	16	AAO87587	DNA encoding Leuco
	21	96	4.0	3399	17	AAO5868	Chicken leucocytoz
	22	83.2	3.4	6644	20	AA33181	Base sequence of t
	23	83.2	3.4	7372	20	AA33182	Base sequence of t
	24	83.2	3.4	7797	20	AA33180	Cowpox virus bsr f
	25	83.2	3.4	7996	20	AA33185	Base sequence of t
	26	82.6	3.4	5940	21	AA70105	Plasmodium falcipa
	27	82	3.4	1998	21	AA70212	Plasmodium falcipa
	28	80.8	3.3	4590	7	AA60472	Sequence encoding
	29	75.2	3.1	3579	21	AA70099	Plasmodium falcipa
	30	74.6	3.1	3549	21	AA70223	Plasmodium falcipa
	31	74.6	3.1	5340	22	AA26289	P. falciparum eg5
C	32	73.6	3.0	1527	21	AA70121	Plasmodium falcipa
	33	72.4	3.0	3927	21	AA70101	Plasmodium falcipa
	34	72.2	3.0	3567	21	AA70117	Plasmodium falcipa
	35	71	2.9	1035	21	AA70136	Human colon cancer
	36	71	2.9	3744	21	AA70149	Plasmodium falcipa
	37	70.8	2.9	7326	21	AA70095	Plasmodium falcipa
	38	70	2.9	7458	21	AA70106	Plasmodium falcipa
	39	69.2	2.8	10640	22	AAO3729	P. falciparum telo
	40	69	2.8	468	21	AA62126	DNA encoding a cal
	41	69	2.8	4611	21	AA62127	Sequence encoding
C	42	68.4	2.8	1612	7	AA60392	Oligonucleotide D1
	43	67.4	2.8	244	22	AA58238	Plasmodium falcipa
	44	67	2.8	3837	21	AA70211	Plasmodium falcipa
	45	66.8	2.7	5994	21	AA70222	Plasmodium falcipa

ALIGNMENTS

RESULT	1	
ID	AAC86015	standard; cdna; 2430 bp.
AC	AAC86015;	
DT	29-AUG-2001	(first entry)
XX		
DE	Wild type B. anthracis lethal factor coding sequence.	
XX		
KW	Lethal factor; LF; immunogen; LF4; protective antigen; PA; DNA vaccine;	
KW	humoral; cell-mediated; immune memory response; ss.	
XX		
OS	Bacillus anthracis.	
XX		
FH	Key	Location/Qualifiers
FT	sig-peptide	1..99
FT		/*tag= a
FT	mat-peptide	100..2430
FT		/*tag= b
FT		/product= "LF"
FT		124..855
FT		/*tag= c
FT	misc_RNA	/product= "Encodes LF4 peptide"
XX		
XX	WO200145639-A2.	
PN		
XX	28-JUN-2001.	
PD		
XX		
PF	21-DEC-2000; 2000WO-US4912.	
XX		
PR	22-DEC-1999; 99US-0171459.	
XX		
PA	(OHIS) UNIV OHIO STATE RES FOUND.	

PA (GALL/) GALLOWAY D R.
PA (MATE/) MATECZUN A J.
XX Galloway DR, Mateczun AJ;
XX WPI: 2001-408540/43.
DR P-PSDB: AAB47305.
XX
PT protecting animal against lethal infection with *Bacillus anthracis*, by
PT administering wild type or mutated form of *Bacillus anthracis* lethal
PT factor protein or its fragment or a nucleic acid encoding the mutated
PT protein -
XX
PS Disclosure: Fig 1: 33pp: English.
XX
CC This sequence encodes the B. anthracis lethal factor (LF). An
CC immunogenic fragment of LF, LF4, can be used to produce an immune
CC response which protects an animal against lethal infection with
CC *Bacillus anthracis*. DNA encoding the B. anthracis LF can be used
CC in conjunction with DNA encoding the protective antigen (PA) in a
CC DNA vaccine. Using a DNA vaccine which encodes the mutated LF protein
CC or fragment alone or in combination with a DNA encoding the PA protein
CC or its fragment, both components (humoral and cell-mediated) of the
CC immune system are stimulated, which results in longer term immune
CC memory response. The combined use of a mutated LF and PA gene or their
CC fragments results in a higher level of immune response, as judged by
CC overall serum antibody titers for LF and PA antigens, than the use of
CC either LF or PA genes in separate immunizations.
XX
SQ Sequence 2430 BP: 985 A; 296 C; 445 G; 704 T; 0 other;

Query Match 100.0%; Score 2430; DB 22; Length 2430;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 2430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaataataaaagaattataaaagaatgaatgaatgctgtttagtaacagaatt 60
DB 1 atgaatataaaagaattataaaagaatgaatgaatgctgtttagtaacagaatt 60
QY 61 acttgagtgctccgctcttaccctgtacagggcgcgcgatgtgtatgta 120
DB 61 acttgagtgctccgctcttaccctgtacagggcgcgcgatgtgtatgta 120
QY 121 ggtatgcacgtataaagaagaagaataaagaatgaatgaatgaaga 180
DB 121 ggtatgcacgtataaagaagaagaataaagaatgaatgaatgaaga 180
QY 181 cgaataaacaacaggaagaacatttaagaagaatcaatgaacaactgttaaaataga 240
DB 181 cgaataaacaacaggaagaacatttaagaagaatcaatgaacaactgttaaaataga 240
QY 241 gtaaaaggagggaagctgttaaaaggagcagcagaagaactacttgaagaatgaca 300
DB 241 gtaaaaggagggaagctgttaaaaggagcagcagaagaactacttgaagaatgaca 300
QY 301 tctgtgcttttagagatgtataaagaatgtgagaagaatataatgtgtgtat 360
DB 301 tctgtgcttttagagatgtataaagaatgtgagaagaatataatgtgtgtat 360
QY 361 attcaaaaacatatcttttagaagcatatcgtgaagaataaagaagaatgacatt 420
DB 361 attcaaaaacatatcttttagaagcatatcgtgaagaataaagaagaatgacatt 420
QY 421 tatgggaagaatgctttattacatgaacattatgtatcgtcaaaagaagaatgaccc 480
DB 421 tatgggaagaatgctttattacatgaacattatgtatcgtcaaaagaagaatgaccc 480
QY 481 gtaactgttaacaaatcttcggaagaatattgtagaanaatactgaaaggcactgaaagt 540
DB 481 gtaactgttaacaaatcttcggaagaatattgtagaanaatactgaaaggcactgaaagt 540
QY 541 tatattgaataatgtaagatatatacaaggatatattttaagtaaaatatacaacatat 600

DB 541 tatattgaataatgtaagatatatacaaggatatattttaagtaaaatatacaacatat 600
QY 601 cagaattctttagatgtataataatcacattaaagaatgctgattcagaatgacaagt 660
DB 601 cagaattctttagatgtataataatcacattaaagaatgctgattcagaatgacaagt 660
QY 661 ctttaattactaatcagcttaaggacatccacagacttctgtgaatctgtgaa 720
DB 661 ctttaattactaatcagcttaaggacatccacagacttctgtgaatctgtgaa 720
QY 721 caaaatagcaatgaggtataaagaatgttgcgaagaagctttgcataatatcgagcca 780
DB 721 caaaatagcaatgaggtataaagaatgttgcgaagaagctttgcataatatcgagcca 780
QY 781 cagcatcgatgctttttagacgttttagcccggaagcttttaattacatgataaatt 840
DB 781 cagcatcgatgctttttagacgttttagcccggaagcttttaattacatgataaatt 840
QY 841 aacgaacaagaataaatatctatccttgaagaactaaagaatcaacagatgctgtcaaga 900
DB 841 aacgaacaagaataaatatctatccttgaagaactaaagaatcaacagatgctgtcaaga 900
QY 901 tatgaaaaatgggaaaaagataaagaacgcctatcaacactgtagagatctttatctgaa 960
DB 901 tatgaaaaatgggaaaaagataaagaacgcctatcaacactgtagagatctttatctgaa 960
QY 961 gaaagaaggagcttttaaaagaagctgcagatcccatggaacaaagaatgacata 1020
DB 961 gaaagaaggagcttttaaaagaagctgcagatcccatggaacaaagaatgacata 1020
QY 1021 attcatctttatctcaagaagaagaagcttctaaagaatacaaatgataatgtagt 1080
DB 1021 attcatctttatctcaagaagaagaagcttctaaagaatacaaatgataatgtagt 1080
QY 1081 gattttttatctacgtgagaagaagaagcttttaaaagaactcaaatgatatctgtgatt 1140
DB 1081 gattttttatctacgtgagaagaagaagcttttaaaagaactcaaatgatatctgtgatt 1140
QY 1141 tctttatctgaagaagaagaagaagcttttaaaatagaatcaggtgtagtataaccc 1200
DB 1141 tctttatctgaagaagaagaagaagcttttaaaatagaatcaggtgtagtataaccc 1200
QY 1201 ttatctgaaaaagaagaagaagcttttaaaagaagctgaaactgataatcaacatatgatt 1260
DB 1201 ttatctgaaaaagaagaagaagcttttaaaagaagctgaaactgataatcaacatatgatt 1260
QY 1261 attaatcaaaaggttgcagaatatacagaagggttaattgtagtccgtaaatcaattatgatt 1320
DB 1261 attaatcaaaaggttgcagaatatacagaagggttaattgtagtccgtaaatcaattatgatt 1320
QY 1321 gtaagaagaacgatatataaaggatattcaaaaatattgtagctttatatacaatcatt 1380
DB 1321 gtaagaagaacgatatataaaggatattcaaaaatattgtagctttatatacaatcatt 1380
QY 1381 ggaagtaacctgtacataaanaatttatctgtatgaaataatgaatacaataaacttaca 1440
DB 1381 ggaagtaacctgtacataaanaatttatctgtatgaaataatgaatacaataaacttaca 1440
QY 1441 ggaacccctagtgcgatttagttatccactgataataacaaatataatagagatt 1500
DB 1441 ggaacccctagtgcgatttagttatccactgataataacaaatataatagagatt 1500
QY 1501 ttcaatgaaatcaaaaaaatttcaaatatgatttccagtaactatgattgtgtat 1560
DB 1501 ttcaatgaaatcaaaaaaatttcaaatatgatttccagtaactatgattgtgtat 1560
QY 1561 ataaatgaaaggccttgatagataatgagcgtttgaaatggaagaatccaattatcacca 1620
DB 1561 ataaatgaaaggccttgatagataatgagcgtttgaaatggaagaatccaattatcacca 1620
QY 1621 gatactcgagcggatatttaagaagaatgaaagcttatattcaagaagaactcgttcg 1680

Db 961 gtaactgtatcccaatcttcggaagatctgttagaanaatacgaagaagcactggaacgtc 1020
 Oy 541 tatatgaataagtagaagatatalcaagaaggaattttaagtaaaaataatcaacacata 600
 Db 1021 tatatgaataagtagaagatatalcaagaaggaattttaagtaaaaataatcaacacata 1080
 Oy 601 cagaatatttttagatgatatataaccatataaaatgcacatcgatccagatggaagaat 660
 Db 1081 cagaatatttttagatgatatataaccatataaaatgcacatcgatccagatggaagaat 1140
 Oy 661 ctttatattactaatacgaatttaaggaaacatccacagacatttcgttaaatcttgga 720
 Db 1141 ctttatattactaatacgaatttaaggaaacatccacagacatttcgttagaattcttgga 1200
 Oy 721 caaaatagcaatgagtagaagaagatatttgcgaagccttgcataatatacgaagca 780
 Db 1201 caaaatagcaatgagtagaagaagatatttgcgaagccttgcataatatacgaagca 1260
 Oy 781 cagcatcgatgattttaagctttatcgacccggaagcctttaataatcaatgataaatt 840
 Db 1261 cagcatcgatgattttaagctttatcgacccggaagcctttaataatcaatgataaatt 1320
 Oy 841 aacgaacaagaataaatcatccctgtgagaagacttaagaatcaacggaatgctgtcaaga 900
 Db 1321 aacgaacaagaataaatcatccctgtgagaagacttaagaatcaacggaatgctgtcaaga 1380
 Oy 901 tatgaaaaatgaggaaaaagataaaacagcatcaacactggaagcgattccttatcgaa 960
 Db 1381 tatgaaaaatgaggaaaaagataaaacagcatcaacactggaagcgattccttatcgaa 1440
 Oy 961 gaaggaagaagacctttaaaaaagctgcagatcccatcgagccaaagaagaatgacata 1020
 Db 1441 gaaggaagaagacctttaaaaaagctgcagatcccatcgagccaaagaagaatgacata 1500
 Oy 1021 attcatcttctatctcaagaagaagaagccttcaaaaaagatacaaaatgataagtagt 1080
 Db 1501 attcatcttctatctcaagaagaagaagccttcaaaaaagatacaaaatgataagtagt 1560
 Oy 1081 gattttttatctcaactggaagaagaagcttttaaaaaagctcaaatgataatctcgtagt 1140
 Db 1561 gattttttatctcaactggaagaagaagcttttaaaaaagctcaaatgataatctcgtagt 1620
 Oy 1141 tctttatctgaaagaagaagaagcctttaaaatagaatcaagtgagtagttagtccct 1200
 Db 1621 tctttatctgaaagaagaagaagcctttaaaatagaatcaagtgagtagttagtccct 1680
 Oy 1201 ttaactgaaaaaagaagaagcttttaaaaaagctgaaactggaattcaacacataagct 1260
 Db 1681 ttaactgaaaaaagaagaagcttttaaaaaagctgaaactggaattcaacacataagct 1740
 Oy 1261 attaatcaaaagctgcagaatcaagaagaaggttaattgatagtccgtcaaatcttgat 1320
 Db 1741 attaatcaaaagctgcagaatcaagaagaaggttaattgatagtccgtcaaatcttgat 1800
 Oy 1321 gtaagaagaagcatcaaaagggatattcaaaatattgagctgttaatacaatccatt 1380
 Db 1801 gtaagaagaagcatcaaaagggatattcaaaatattgagctgttaatacaatccatt 1860
 Oy 1381 ggaagtaacctgttacaataaataattattgttatagaaataatgataatcaacttaca 1440
 Db 1861 ggaagtaacctgttacaataaataattattgttatagaaataatgataatcaacttaca 1920
 Oy 1441 gcaaccccttagtgcgagatttagtattccactgataataactcaaaatgaaggttatt 1500
 Db 1921 gcaaccccttagtgcgagatttagtattccactgataataactcaaaatgaaggttatt 1980
 Oy 1501 ttcaatgaaattcaaaaaaattcaaaatataagattttcagtaagactataatgttgtt 1560
 Db 1981 ttcaatgaaattcaaaaaaattcaaaatataagattttcagtaagactataatgttgtt 2040
 Oy 1561 ataatagaaagcctgcatatagataatgagcgttgaaatgagagatccaatatcacca 1620

Db 2041 ataataagaaagcctgcatatagataatgagcgtttgaaatggaagatccaatatcacca 2100
 Oy 1621 gatactcgagcggagatatttagaanaaaatggaagccttatalatcaaaagaacatcggtcgt 1680
 Db 2101 gatactcgagcggagatatttagaanaaaatggaagccttatalatcaaaagaacatcggtcgt 2160
 Oy 1681 gaataaagatgatacaaaatataagcaatccgaaagaatataatgaagattgagcgt 1740
 Db 2161 gaataaagatgatacaaaatataagcaatccgaaagaatataatgaagattgagcgt 2220
 Oy 1741 aaagtatgccaagaagtagaataatagatacaaaatccagaagcacaagttaaatataat 1800
 Db 2221 aaagtatgccaagaagtagaataatagatacaaaatccagaagcacaagttaaatataat 2280
 Oy 1801 caggaaatggaataaagcatlaaggttaccaaaatatacaaaacttatacatcaacgtg 1860
 Db 2281 caggaaatggaataaagcatlaaggttaccaaaatatacaaaacttatacatcaacgtg 2340
 Oy 1861 cataatagatattgcatccaatattgttagaagaagtctattataatgaaatgaa 1920
 Db 2341 cataatagatattgcatccaatattgttagaagaagtctattataatgaaatgaa 2400
 Oy 1921 aataatattcaaaagtattcttaaaaaaggtcaacaattacttagttagttaga 1980
 Db 2401 aataatattcaaaagtattcttaaaaaaggtcaacaattacttagttagttaga 2460
 Oy 1981 agattgtttttaccgatatattcctccaaatataagctgaaacaatatacacatcaagat 2040
 Db 2461 agattgtttttaccgatatattcctccaaatataagctgaaacaatatacacatcaagat 2520
 Oy 2041 gagaataagaaagattatattcaaaaggttatatggtccgaagatcccggttctatata 2100
 Db 2521 gagaataagaaagattatattcaaaaggttatatggtccgaagatcccggttctatata 2580
 Oy 2101 ctcacatgaccttcaaaaaggtgtagaatcaaggaatgagtagtgaaggtttataacagaa 2160
 Db 2581 ctcacatgaccttcaaaaaggtgtagaatcaaggaatgagtagtgaaggtttataacagaa 2640
 Oy 2161 ttgtgacatgctgtgtagatattctgtagatatctatttagataagaaccaatctgattta 2220
 Db 2641 ttgtgacatgctgtgtagatattctgtagatatctatttagataagaaccaatctgattta 2700
 Oy 2221 gttacaattctcaaaaatctatgatatattttaagaagaaggtgtagtatttaactctg 2280
 Db 2701 gttacaattctcaaaaatctatgatatattttaagaagaaggtgtagtatttaactctg 2760
 Oy 2281 tatggaagaacaatgaaagcgaatatttttcgaagaagcctttaggtttaaagctatcag 2340
 Db 2761 tatggaagaacaatgaaagcgaatatttttcgaagaagcctttaggtttaaagctatcag 2820
 Oy 2341 gaccatgctgaaagcttttaaaagttcaaaaaaatgctcgaaaaacttccaatttaaac 2400
 Db 2821 gaccatgctgaaagcttttaaaagttcaaaaaaatgctcgaaaaacttccaatttaaac 2880
 Oy 2401 gatcagattaaattcaattatcaatcaaa 2430
 Db 2881 gatcagattaaattcaattatcaatcaaa 2910

RESULT 3
 AAS00250
 ID AAS00250 standard; DNA; 1455 BP.
 XX AAS00250;
 AC AAS00250;
 XX AAS00250;
 DR 31-MAY-2001 (first entry)
 XX
 DE LFn-Bcl-Xl apoptosis-modifying fusion protein, DNA sequence.
 XX
 XX Human; LFn-Bcl-Xl; apoptosis; cancer; spinal muscular atrophy; ds;
 KW anthrax lethal factor; neoplasm; tumour; hyper-proliferation;
 KW Alzheimer's disease; neurodegenerative disorder; stroke;
 KW transient ischaemic neuronal injury; spinal cord injury;

KW Huntington's disease.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Corynebacterium diphtheriae.
OS Chimeric - Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..1455
FT /tag= a
FT /partial
FT /product= "Lfn-Bcl-Xl fusion protein"
FT /note= "LF is anthrax lethal factor. No stop codon given"
FT misc-feature 13..30
FT /tag= b
FT /note= "6x Histidine tag"
FT misc-feature 61..828
FT /tag= c
FT /note= "Lethal factor codons 1 to 255"
FT misc-feature 829..1455
FT /tag= d
FT /note= "Bcl-Xl codons 1 to 209"
XX
PN WO200112661-A2.
XX
PD 22-FEB-2001.
XX
PF 15-AUG-2000; 2000MO-US22293.
XX
PR 16-AUG-1999; 990US-0149220.
XX
PA (HAND) HARVARD COLLEGE.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Youle RJ, Liu X, Collier RJ;
XX
DR WPI. 2001-218343/22.
XX
P-PSDB: AAU00222.
XX
PT Novel fusion protein for modifying apoptosis in target cell and
PT reducing apoptosis after transient ischaemic neuronal injury, has two
PT domains which targets protein to a cell and modifies apoptotic response
PT of cell -
XX
PS Claim 5; Page 61-63; 65pp; English.
XX
XX The sequence represents the coding sequence of Lfn-Bcl-Xl apoptosis-
XX modifying fusion protein comprising anthrax lethal factor (Lf) sequence
XX fused to Bcl-Xl. The functional apoptosis-modifying fusion protein is
XX capable of binding a target cell and integrating into or crossing a
XX cellular membrane of the target cell. The apoptosis-modifying fusion
XX protein comprises at least two domains: the DTR domain, which targets the
XX fusion protein to the target cell and the Bcl-Xl domain, which modifies
XX an apoptotic response of the target cell. The fusion protein is useful
XX for modifying (inhibiting or enhancing) apoptosis in a target cell, such
XX as neuron, lymphocyte, cancer, neoplasm, macrophage, epithelial, stem,
XX tumour or hyper-proliferative cell or an adipocyte. It is also useful for
XX reducing apoptosis in a subject after transient ischaemic neuronal
XX injury, especially spinal cord injury. The fusion protein may be used to
XX treat various diseases and injury conditions through inhibition or
XX enhancement of apoptotic cellular response, including neurodegenerative
XX disorders such as Alzheimer's disease, Huntington's disease, spinal
XX muscular atrophy, stroke episodes and unregulated cell growth as in
XX tumours and various cancers. The apoptosis-modifying fusion protein can
XX be delivered effectively throughout the body and targeted to selective
XX tissue and cells.
XX
SQ Sequence 1455 BP; 487 A; 266 C; 369 G; 333 T; 0 other;

Query Match 31.6%; Score 767.6; DB 22; Length 1455;
Best Local Similarity 99.5%; Pred. No. 1e-124;
Matches 770; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 94 caggggcgggcggtcattggtgattgattgacgttaaagaagaaagaaataaa 153
|||
Db 58 catatggcgggcggtcattggtgattgattgacgttaaagaagaaagaaataaa 117
QY 154 gatgagataaagaagaaagaatgaagaacgaataaaacacaggaagagcattaaagaa 213
|||||
Db 118 gatgagataaagaagaaagaatgaagaacgaataaaacacaggaagagcattaaagaa 177
QY 214 atcatgaacaacattgtaaaatagaagaataaaggaggagagctgttaaaagagca 273
|||||
Db 178 atcatgaacaacattgtaaaatagaagaataaaggaggagagctgttaaaagagca 237
QY 274 gcagaaagcttacttggaagaagfaccctctgattgtttaaggtgttaagaagcaattgga 333
|||||
Db 238 gcagaaagcttacttggaagaagfaccctctgattgtttaaggtgttaagaagcaattgga 297
QY 334 ggaagataatataattggtgattggtatatacaaaacataatcttaagaagcattcc 393
|||||
Db 298 ggaagataatataattggtgattggtatatacaaaacataatcttaagaagcattcc 357
QY 394 gaagataaagaagaaataaagaacattatgggaagaatgcttattatgaacattatc 453
|||||
Db 358 gaagataaagaagaaataaagaacattatgggaagaatgcttattatgaacattatc 417
QY 454 gtatagcaaaagaagataatgaacccgtactgttaattcaatcttcggaagattatgta 513
|||||
Db 418 gtatagcaaaagaagataatgaacccgtactgttaattcaatcttcggaagattatgta 477
QY 514 gaaataactgaaagagcagcagcgttattatgaataagataagataatcaagagat 573
|||||
Db 478 gaaataactgaaagagcagcagcgttattatgaataagataagataatcaagagat 537
QY 574 atttaagtaaaatgaataacacataatcaagaatttttagatgtataataccataaa 633
|||||
Db 538 atttaagtaaaatgaataacacataatcaagaatttttagatgtataataccataaa 597
QY 634 aatgcattgattcagaatgacaagaatcttttattactaatccttaaggagacatccc 693
|||||
Db 598 aatgcattgattcagaatgacaagaatcttttattactaatccttaaggagacatccc 657
QY 694 acagactttctgtagaattcttcggaacaaatagaagaatgaggtacagaagaatttcg 753
|||||
Db 658 acagactttctgtagaattcttcggaacaaatagaagaatgaggtacagaagaatttcg 717
QY 754 aaagctttgcataattatcagagccacagacatcgatgtttttacagctttatgcacg 813
|||||
Db 718 aaagctttgcataattatcagagccacagacatcgatgtttttacagctttatgcacg 777
QY 814 gaagctttaattacatgataaatttaacgaaacaaagaataatctatccttg 867
|||||
Db 778 gaagctttaattacatgataaatttaacgaaacaaagaataatctatccttg 831
RESULT 4
AAQ70181
ID AAQ70181 standard; DNA; 1368 BP.
XX
AC AAQ70181;
XX
DT 03-APR-1995 (first entry)
XX
DE LF(1-254)--TR--PE(401-602) toxin fusion protein coding sequence.
XX
KW Anthrax; Bacillus anthracis; fusion protein; lethal factor;
KW protective antigen; cell killing; targeting; pathogen;
KW intracellular; HIV; human immunodeficiency virus; toxin;
KW Pseudomonas; exotoxin; ss.
XX
OS Bacillus anthracis.
OS Pseudomonas sp.
XX
FH Key Location/Qualifiers
FT CDS 1..1368

```
FT      /*tag= a
XX      /product= Anthrax toxin fusion protein.
PN
XX      MO9418332-A.
PD      18-AUG-1994.
XX
XX      14-FEB-1994; 94MO-US01624.
XX
XX      12-FEB-1993; 93US-0021601.
PR      25-JUN-1993; 93US-0082849.
XX
PA      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI      Arora N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;
XX
XX      WPI: 1994-279753/34.
DR      P-PSDB; AAR60180.
XX
XX      Nucleic acid encoding anthrax toxin fusion protein - useful for
PT      targeting toxin to specific cells, eg for killing tumour cells
PT      or HIV-infected cells
PS
XX      Claim 6; Page 84-86; 124pp; English.
XX
CC      This sequence encodes a fusion protein comprising amino acid
CC      residues 1-254 of the anthrax protective antigen binding domain of
CC      the native anthrax lethal factor, a two residue linker and a
CC      sequence encoding residues 401-602 of a Pseudomonas exotoxin A
CC      activity inducing domain of a second protein. Such fusion proteins
CC      may be useful for the specific killing of tumour cells or the
CC      killing of cells infected with intracellular pathogens, especially
CC      HIV, depending on their components.
XX
SQ      Sequence 1368 BP; 411 A; 316 C; 358 G; 283 T; 0 other;

Query Match      31.4%; Score 762.4; DB 15; Length 1368;
Best Local Similarity 99.2%; Pred. No. 8.1e-124;
Matches 766; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      100 gcgggcggtcgtggtgagtgatgacgtaagaagaagaagaagaagaagaagaatgag 159
D      1 gcgggcggtcgtggtgagtgatgacgtaagaagaagaagaagaagaagaatgag 60
QY      160 aataagagaaaaagatgaaagacgaataaacaacagagaagcatttaagaatcatg 219
D      61 aataagagaaaaagatgaaagacgaataaacaacagagaagcatttaagaatcatg 120
QY      220 aaacacattgtaaaaataagaagtaaaaggaggaagctgttaaaaaaaggagcagagaa 279
D      121 aaacacattgtaaaaataagaagtaaaaggaggaagctgttaaaaaaaggagcagagaa 180
QY      280 aagctacttggaagaagtaaccatctgatttttagagatgtataaagcaatttgagaagaag 339
D      181 aagctacttggaagaagtaaccatctgatttttagagatgtataaagcaatttgagaagaag 240
QY      340 atatatatttgatggtgataataacaaacatatattcttgaagcattatctgaagat 399
D      241 atatatatttgatggtgataataacaaacatatattcttgaagcattatctgaagat 300
QY      400 aagaaaaaataaagaacattatgaggaagaatgctttattatcagaacattatgtatat 459
D      301 aagaaaaaataaagaacattatgaggaagaatgctttattatcagaacattatgtatat 360
QY      460 gcaaaagaagatatgaaccgctacttgtaaccatctcggagaagattatgaagaat 519
D      361 gcaaaagaagatatgaaccgctacttgtaaccatctcggagaagattatgaagaat 420
QY      520 actgaaaagagcactgaaacttattatgaagaataggttaagataatcaaggagataat 579
D      421 actgaaaagagcactgaaacttattatgaagaataggttaagataatcaaggagataat 480
```

```
QY      580 agtaaatatcaaccatatacagaatttttagatgtatataaataaccattaaatgca 639
D      481 agtaaatatcaaccatatacagaatttttagatgtatataaataaccattaaatgca 540
QY      640 tctgattcagatggacaagaatctttatttaactaactagcttaagaacatccccaagac 699
D      541 tctgattcagatggacaagaatctttatttaactaactagcttaagaacatccccaagac 600
QY      700 ttctctgagaattcttggacaataagcaatgagtgatagaagaatatttgcgaagaact 759
D      601 ttctctgagaattcttggacaataagcaatgagtgatagaagaatatttgcgaagaact 660
QY      760 ttgcattattatatacagccacagcatcgtgatttttaacagcttattgacccgaagct 819
D      661 ttgcattattatatacagccacagcatcgtgatttttaacagcttattgacccgaagct 720
QY      820 ttaattacatgagtaaatttaacgaacaagaataaatactatcctcttggaag 871
D      721 ttaattacatgagtaaatttaacgaacaagaataaatactatcctctggaag 772

RESULT      5
AAQ70182
ID      AAQ70182 standard; DNA: 1425 BP.
XX
AC      AAQ70182;
XX
DR      04-APR-1995 (first entry)
XX
DE      LF(1-254)--TR--PE(398-613) toxin fusion protein coding sequence.
XX
KW      Anthrax; Bacillus anthracis; fusion protein; lethal factor;
KW      protective antigen; cell killing; targeting; pathogen;
KW      intracellular; HIV; human immunodeficiency virus; toxin;
KW      Pseudomonas; exotoxin; ss.
XX
OS      Bacillus anthracis.
XX
XX      Pseudomonas sp.
XX
Key      Location/Qualifiers
FT      CDS      1..1416
FT      /*tag= a
FT      /product= Anthrax toxin fusion protein.
XX
XX      MO9418332-A.
XX
XX      18-AUG-1994.
XX
XX      14-FEB-1994; 94MO-US01624.
XX
XX      12-FEB-1993; 93US-0021601.
XX      25-JUN-1993; 93US-0082849.
XX
PA      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI      Arora N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;
XX
XX      WPI: 1994-279753/34.
DR      P-PSDB; AAR60181.
XX
XX      Nucleic acid encoding anthrax toxin fusion protein - useful for
PT      targeting toxin to specific cells, eg for killing tumour cells
PT      or HIV-infected cells
PS
XX      Example 2; Page 87-90; 124pp; English.
XX
CC      This sequence encodes a fusion protein comprising amino acid
CC      residues 1-254 of the anthrax protective antigen binding domain of
CC      the native anthrax lethal factor, a two residue linker and a
CC      sequence encoding residues 398-613 of a Pseudomonas exotoxin A
CC      activity inducing domain of a second protein. Such fusion proteins
CC      may be useful for the specific killing of tumour cells or the
CC      killing of cells infected with intracellular pathogens, especially
```


QY	460	gcaaaagaaggaatgtgaaccgctgactctgttaaccatcttcggaagatattgtagaanaat	519
Db	361	gcaaaagaaggaatgtgaaccgctgactctgttaaccatcttcggaagatattgtagaanaat	420
QY	520	actgaaagaagcgtgaacgctttaaattgaataagtaagataatacaaggaatttta	579
Db	421	actgaaagaagcgtgaacgctttaaattgaataagtaagataatacaaggaatttta	480
QY	560	agtaaatataatcaacatatcagaaatttttagatgatatataaccatataaatatgca	639
Db	481	agtaaatataatcaacatatcagaaatttttagatgatatataaccatataaatatgca	540
QY	640	tcgtatcagaatggaacaagatctttatttaccatcagcttaagaagaatccccaagac	699
Db	541	tcgtatcagaatggaacaagatctttatttaccatcagcttaagaagaatccccaagac	600
QY	700	ttttctgtagaattctcttggaacaataatagcaatgagtaagtaacaaagtaatttgcaagct	759
Db	601	ttttctgtagaattctcttggaacaataatagcaatgagtaagtaacaaagtaatttgcaagct	660
QY	760	tttgcataataatcagcgacacagcatcgtgatgtttacaagcttatacgaccggaagct	819
Db	661	tttgcataataatcagcgacacagcatcgtgatgtttacaagcttatacgaccggaagct	720
QY	820	tttaattacatcgtgataaatttaaacygaacaagaataataactatc	863
Db	721	tttaattacatcgtgataaatttaaacygaacaagaataataactatc	764
RESULT 7			
AF58252/c	AAF58252/c		
ID	AAF58252 standard; DNA; 936 BP.		
AC	AAF58252;		
XX	24-APR-2001 (first entry)		
DT	Oligonucleotide D1835.		
XX	Electron-transfer group; ETM; mismatch; genotyping;		
KM	gene expression; ss.		
OS	Synthetic.		
XX	MO200107665-A2.		
PN	01-FEB-2001.		
XX	26-JUL-2000; 2000MO-US20476.		
PF	26-JUL-1999; 99US-0145695.		
XX	17-MAR-2000; 2000US-0190259.		
PR	(CLIN-) CLINICAL MICRO SENSORS INC.		
XX	Umek RM;		
PI	WPI; 2001-159728/16.		
DR	Nucleic acids containing electron-transfer group, useful as labels in		
XX	hybridization assays, e.g. for genotyping, allowing repeat analyses on		
PT	a single surface		
XX	Example 6; Page 127; 159pp; English.		
XX	The present invention relates to a composition comprising two nucleic		
CC	acids each containing an electron-transfer group (ETM) having		
CC	different redox potentials. The invention is used for electronic		
CC	detection of nucleic acids, especially of substitutions (mismatches)		
CC	and single-nucleotide polymorphisms, e.g. for genotyping,		
XX	monitoring gene expression.		

```

50 Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other:
Query Match      8.7%, Score 211.2; DB 22; Length 936;
Best Local Similarity 1.1%; Pred. No. 3,1e-28;
Matches 9; Conservative 559; Mismatches 222; Indels 0; Gaps 0.

yy 1182 ggtgatgatgtgtaatccttctatcgcgaaagaagaaggttttcaaaaagctaaact 1241
    ||| : : : : : : : : : : : : : : : : : : : : : : : : :
pb 793 GGGGGGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 734

yy 1242 tgcataccaacatagatgatataatcaaaaggtgcgaagatcaggaggttaattgatag 1301
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
pb 733 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 674

yy 1302 tcgcgaatatacttgatgtaagaagcagatataaaaggagatccaatatgtatgc 1361
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
pb 673 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 614

yy 1362 ttattacaatccaatcgtgaagtagaactgttacataaaatttcttgatgtaaaatc 1421
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
pb 613 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 554

yy 1422 gaataccaataacctacagcaacctaggctgcgatttagttgattccaactgataatc 1481
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
pb 553 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 494

yy 1482 taagaatagaggtatttccaatgaaatccaataaaattccaatagatattctag 1541
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
pb 493 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 434

yy 1542 taactaatgatgttctgatacaatgaaagcctcattagataatgagcgtttgaaatg 1601
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
pb 433 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 374

yy 1602 gagaatccaattacaccagatactcgagcagatattagaaaatggaagcttatatt 1661
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
pb 373 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 314

yy 1662 acaagaacaatcgcttcggaataaagatgtaacaaatataagcaatccgaaaaga 1721
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
pb 313 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 254

yy 1722 atataaagatctgacgaagtagtgcgaagaagtaaaatagatacaaaaatccaaga 1781
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
pb 253 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 194

yy 1782 agcacagttaaataaataatcggaatggaataaagcattaggttaccacaaatatacaaa 1841
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
pb 193 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 134

yy 1842 gcttatcatccaagtcgataatagatatgcatccaatattgtaagaagtcgatatt 1901
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
pb 133 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 74

yy 1902 aatattgaatgaatggaataaataatataccaagtgatcttataaaaagtaacaatta 1961
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
pb 73 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 14

yy 1962 cttagttgat 1971
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
pb 13 WWWWWWWWWWWWW 4

RESULT      8
AAFS8254/C
ID      AAF58254 standard; DNA; 936 BP.
XX      AAF58254;
XX
XX      24-APR-2001 (first entry)
XX
DE      Oligonucleotide D1875.

```

```

XX Electron-transfer group; ETW; mismatch; genotyping;
KW gene expression; ss.
XX Synthetic.
XX WO200107665-A2.
XX
XX 01-FEB-2001.
XX
XX 26-JUL-2000; 2000WO-US20476.
XX
XX 26-JUL-1999; 99US-0145695.
XX
XX 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX Umek RM;
XX
XX WPI; 2001-159728/16.
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface
XX
XX Example 6; Page 127; 159pp; English.
XX
XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETW) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.
XX
XX Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;
XX
Query Match      8.7%; Score 211.2; DB 22; Length 936;
Best Local Similarity 1.1%; Pred. No. 3.1e-28;
Matches 9; Conservative 559; Mismatches 222; Indels 0; Gaps 0;
OY 1182 ggtgagatagtaaccccttctctcgtgaaagaaagagttttaaaagctgaact 1241
DB 793 GGGGGGCMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 734
OY 1242 tgatatcaacataatgataatcaaaaggtgcagagacagaggttaattgatatag 1301
DB 733 MNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 674
OY 1302 tccgtcaattacttgatgtaagaagcagataaaagagatccaataattgattgc 1361
DB 673 MNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 614
OY 1362 tttaatacaatccatcgtgaagctgtaacataaattatctgtatgataaatat 1421
DB 613 MNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 554
OY 1422 gaatacaataacactacagcaacctaggtgcgattagttgattccaactgataaac 1481
DB 553 MNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 494
OY 1482 taaataatagaggtatttcaatgaattcaaaaaaaattccaatatagattctag 1541
DB 493 MNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 434
OY 1542 taactatgatgtgtgataataatgaagcgctcatagataatgagcgtttgaatg 1601
DB 433 MNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 374
OY 1602 gagaatcaattatcacagataactcagcagagataatttgaataatggaagcttatt 1661
DB 373 MNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 314

```

```

OY 1662 acaagaacatcgctcgtgaaataaagatgtacaaataatgaacatccgaaaaaga 1721
DB 313 MNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 254
OY 1722 atataaaggatgtgcgaagtagtccaaagagtaaaatagatacaaaatccaaga 1781
DB 253 MNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 194
OY 1782 agcacagttaaataataatcagaatgataaagcatltaggttaccacaaataacaa 1841
DB 193 MNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 134
OY 1842 gcttattacatcaacgtgcataatagatagcatccaatattgaaagtgctattt 1901
DB 133 MNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 74
OY 1902 aatatgaatgaatgaaaaataatattcaaaagtgatctataaaaggtaacaaatla 1961
DB 73 MNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 14
OY 1962 cttagttgat 1971
DB 13 MNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4

RESULT 9
AAFS8257/C
ID AAF58257 standard; DNA; 936 BP.
XX
XX AAF58257;
AC XX
XX 24-APR-2001 (first entry)
DT XX
XX Oligonucleotide D1954.
DE XX
XX Electron-transfer group; ETW; mismatch; genotyping;
KW gene expression; ss.
XX
XX Synthetic.
XX
XX WO200107665-A2.
XX
XX 01-FEB-2001.
XX
XX 26-JUL-2000; 2000WO-US20476.
XX
XX 26-JUL-1999; 99US-0145695.
XX
XX 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX Umek RM;
XX
XX WPI; 2001-159728/16.
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface
XX
XX Example 6; Page 127; 159pp; English.
XX
XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETW) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.
XX
XX Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;
XX
Query Match      8.7%; Score 211.2; DB 22; Length 936;
Best Local Similarity 1.1%; Pred. No. 3.1e-28;

```



```

RESULT 11
AAFS8262/c
ID AAF58262 standard; DNA; 936 BP.
XX
XX
AC AAF58262:
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2007.
XX
KM Electron transfer group; ETM; mismatch; genotyping;
KM gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000MO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI timek RM:
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface.
XX
XX
PS Example 6; Page 128; 159pp; English.
XX
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;
XX
Query Match 8.7%; Score 211.2; DB 22; Length 936;
Best Local Similarity 1.1%; Pred. No. 3.1e-28;
Matches 9; Conservative 559; Mismatches 222; Indels 0; Gaps
OY 1182 ggtgtgatagtacgttaactcttactctgaaaaagaagaagatttttaaaaaagctgaact 1241
DB 793 GGGGGGCGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 734

```

RESULT	12
AAF58255/c	
ID	AAF58255 standard; DNA, 938 BP.
XX	
AC	AAF58255;
XX	
DT	24-APR-2001 (first entry)
XX	
DE	Oligonucleotide D1876.
XX	
KW	Electron-transfer group; ETM, mismatch; genotyping;
KW	gene expression; ss.
XX	
OS	Synthetic.
XX	
PN	W0200107665-A2.
XX	
PD	01-FEB-2001.
XX	

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 2, 2001, 13:49:05 ; Search time 1758.69 Seconds
(without alignments)
22794.322 Million cell updates/sec

Title: US-09-747-521-1
2430
Sequence: 1 atgataataaaagaatt.....agtcattataactcataa 2430

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_hig:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_ov:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_higo_hum:*
31: em_higo_inv:*
32: em_higo_rod:*
33: em_hig_hum:*
34: em_hig_inv:*
35: em_hig_rod:*
36: em_hig_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2430	100.0	3291	1	BACIEFB	M29081 Bacillus an
2	2430	100.0	3291	6	I33395	I33395 Sequence 1
3	2430	100.0	3291	6	I69373	I69373 Sequence 1
4	2430	100.0	3631	1	BACIEFB	M30210 B.anthraxis
5	2430	100.0	181654	1	AF065404	AF065404 Bacillus
6	767.6	31.6	1455	6	AX085496	AX085496 Sequence
7	762.4	31.4	1368	6	I69375	I69375 Sequence 5
8	762.4	31.4	1368	6	I69375	I69375 Sequence 5
9	762.4	31.4	1425	6	I33398	I33398 Sequence 7
10	762.4	31.4	1425	6	I33398	I33398 Sequence 7
11	762.4	31.4	1524	6	I33399	I33399 Sequence 9
12	762.4	31.4	1524	6	I69377	I69377 Sequence 9
13	162.6	6.7	2990	1	BACCYA	M23179 Bacillus an
14	162.6	6.7	2990	6	A02546	A02546 B.anthraxis
15	162.6	6.7	3420	1	BACCYA	M24074 Bacillus an
16	162.6	6.7	181654	1	AF065404	AF065404 Bacillus
17	159.4	6.6	2860	1	AF031382	AF031382 Bacillus
18	158.2	6.5	2990	6	A07289	A07289 B.anthraxis
19	110	4.5	143381	9	AC091214	AC091214 Homo sapi
20	103.4	4.3	158398	2	AC011146	AC011146 Homo sapi
21	103.2	4.2	130540	2	AC079417	AC079417 Mus muscu
22	102.8	4.2	14867	3	AE001398	AE001398 Plasmodu
23	102.2	4.2	4102	3	AF270648	AF270648 Plasmodu
24	101.6	4.2	298469	3	AE003846	AE003846 Drosophi
25	101	4.2	2426	8	SD049822	U49822 Saccharomyc
26	100.2	4.1	155204	2	AC007926	AC007926 Trypanoso
27	98.4	4.0	1686	22	E08995	E08995 DNA encodin
28	98.2	4.0	110000	2	PFMAL1P1_1	Continuation (2 of
29	97	4.0	104992	2	AC005504	AC005504 Plasmodu
30	97	4.0	153477	2	AC006278	AC006278 Plasmodu
31	97	4.0	169546	2	AC004157	AC004157 Plasmodu
32	97	4.0	303091	2	AC084799	AC084799 Mus muscu
33	96.8	4.0	5420	3	AF056936	AF056936 Plasmodu
34	96	4.0	2069	22	E10125	E10125 DNA encodin
35	96	4.0	3399	22	E10126	E10126 DNA encodin
36	95.6	3.9	67970	3	PFMAL1P3	AL031746 Plasmodu
37	94	3.9	4995	3	PFAMESA	M69183 Plasmodu
38	93.6	3.9	54345	3	AC084152	AC084152 Caenorhab
39	93.4	3.8	47573	3	AF030694	AF030694 Plasmodu
40	92.8	3.8	15421	3	PFCOMPIRA	X95275 P.falciparu
41	92.4	3.8	111554	8	AP002460	AP002460 Arabidops
42	91.8	3.8	213530	2	AC090493	AC090493 Mus muscu
43	91.6	3.8	7218	6	I66494	I66494 Sequence 14
44	91.2	3.8	120029	2	HSJ282H10	AL132672 Homo sapi
45	90.4	3.7	183584	9	AC012492	AC012492 Homo sapi

ALIGNMENTS

RESULT	1	LOCUS	BACIEFB	DEFINITION	Bacillus anthracis 3291 bp DNA	VERSION	M29081	ACCESSION	M29081.1	GI:143143	KEYWORDS	anthrax toxin; exotoxin.	SOURCE	Bacillus anthracis (clone: pLF74) DNA.	ORGANISM	Bacillus anthracis	REFERENCE	Braeg, T.S. and Robertson, D.L.	AUTHORS	1 (bases 1 to 3291)	TITLE	Nucleotide sequence and analysis of the lethal factor gene (lef)	JOURNAL	from Bacillus anthracis	MEDLINE	Gene 81 (1), 45-54 (1989)	FEATURES	90034185	location/Qualifiers	1..3291	/organism="Bacillus anthracis"
--------	---	-------	---------	------------	--------------------------------	---------	--------	-----------	----------	-----------	----------	--------------------------	--------	--	----------	--------------------	-----------	---------------------------------	---------	---------------------	-------	--	---------	-------------------------	---------	---------------------------	----------	----------	---------------------	---------	--------------------------------

sig_peptide /db_xref="taxon:1392"
481..579 /clone="pLF74"
/note="lethal factor signal peptide"
CDS 481..2910 /note="lethal factor precursor"
/codon_start=1
/transl_table=11
/protein_id="AAA79216.1"
/db_xref="GI:143144"
/translation="MNIRKEPIKVISMSCLVTATLTSGPVPIPLVQAGAGHDVGMHY
KEKENKDENRKDEBRNKKTOEHLKIMKIVKLENGEAEVKALEAEKLEKVPD
VLEMKATGKIVYVDGDTKHSLEAESEDKRKIKDYGDAVLLHEHLYVAKGEYEP
VLVIOSEEDYVENKALNVYEIGKILSRILSLINOPKPDVLTNTKASDSG
QDLFTNOLKEHPDVSVEPLEONSNEVEAFKAPAFIYPIQRHDVLTQAPAEFNY
MDKNEOINLSLEBKDRLMISREKKEKIKQYQHMDSLSEGRLLKQIPIE
PKDDIHSLSOEKELKRLQIDSSDPLSTEEKLEKLIIDISLSEKELNR
IOWNSNPISREKELKRLQIDSSDPLSTEEKLEKLIIDISLSEKELNR
NIDALHOSIGSTLYNKIYLENNMINNTATGADLVSDTNKIRNGIENEEKNF
KYSISSNMIVDINERPALDNERLKMRIQLSPDRAGLNGKILIRNGIGLEIKDVO
IIRKSEKRYRIDAKVVPKSKIDRKIOEALONINONENKALGPKYTKLITFNVHNK
ASNTVESAYLILNEMKNIOSDLIKKVTNTLVONGRFVPTDITLFIADYTHQDI
YEQVHSGLYVPESRILLGPKSCVELRNDSEGIIEFGHADVIVAGYLDKQSDLI
VTNSKRFIDIFKEGSMULTSYGRITNEAEFFAEARLMSDHERLKVOKNAPKTFQF
INDOKRIINS"
580..2907

mat_peptide /note="lethal factor"
BASE COUNT 1333 a 391 c 564 g 1003 t
ORIGIN

Query Match Best Local Similarity 100.0%; Score 2430; DB 1; Length 3291;
Matches 2430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaataaaaaagaattcataaagaattagatgltcgtttagtaacacgaatt 60
Db 481 ATGAATATAAAAAAGATTATTAAGATATGATGATGATGATGATGATGATGAT 120
QY 61 accttaagtggtccgclttatcccccgtgtaagggcgggcggtcgtgtagtga 120
Db 541 ACTTTAGTGGTCCGCTTTATCCCTGTCACAGGGCGGGCGGTCTGATGATGAT 600
QY 121 ggtatgcgttaaaagaagaagaataaagaatgagaataagaagaataagaaga 180
Db 601 GGTATCAGCTTAAG 660
QY 181 cgaataaaaaacagaagaagacattaaaggaatcataaacactgtttaaaataga 240
Db 661 CCAAAATAAAAACAGGAAAGACATTAAAGAAATCATGAACACATTTGTAATAAGAA 720
QY 241 gtaaaaggagggaagcgtttaaaagaagcagcaagaagctacttgaagaagta 300
Db 721 GTAAGAGGGAGGAGCTGTAAAAAGAGCGACGAGAAACCTACTTGAAGAGTACCA 780
QY 301 tctgtgtttttagagatratataaagaacattggaagaataatatatgtggtggt 360
Db 781 TCTGATGTTTATGATATGATTAAGCAATGAGAGAGAGATATATGATGATGATGAT 840
QY 361 attcaaaaaacatactcttgaagaacatctctgaagaataagaagaataagaacatt 420
Db 841 ATTCAAAAAACATATCTTTGAGAACATTATCTGAAGATTAAGAAAAATAAAGACATT 900
QY 421 taagggaagagtgctttattacatgaacattatgtatatacgcaaaagaaggaatga 480
Db 901 TATGGGAAAGATGCTTTATATCATGAACATTATGATATGCAAAAAGAGATATGAACCC 960
QY 481 gtaacttaaccacactctcggaagatatagtagaataactgaagaagcactgaagct 540
Db 961 GTACTTGTATTCATCTCGGAAGATTATGTAGAAAAATACTGAAAGCGCATGACGTT 1020
QY 541 tatatgaataaggtaagatatatacaagggaattttaagtaaatatacaacatat 600

Db 1021 TATTATGAATAGGTAAAGATATTATCAAGGATATTTTAAGTAATAATCAACCAATAT 1080
QY 601 caagaatttttagatgtatataaataaccataaataatgacatcgtatcagaatgacaagat 660
Db 1081 CAGAAATTTTATGATGTATTAATATACCATTTAAAAATGATCGATTCAGATGACAGAT 1140
QY 661 ctttatttactaatcagcttgaagaagatcccaagacattctcgtagaattcttgtaa 720
Db 1141 CTTTATTTATGATATCAGCTTAAAGAACATCCACAGACTTTTTCGTAAATTCCTGGAA 1200
QY 721 caaaatagcaatgaggtacagaagaatatttgcgaagccttgacatatatctcgaaga 780
Db 1201 CAAAATAGCAATGAGGTACAAAGATTTGGCAAGCTTTTGCAATATATATTCAGACGA 1260
QY 781 caagatcgtgagtgctttacagcttatacgaccggaagcctttaatataatgaataatt 840
Db 1261 CAGCATCGTGATGTTTATCAGCTTATTCACCGGAAAGCTTTTAATATGATGATTAATTT 1320
QY 841 aacgacacagaataaatactatcccttggagaagaacttaagaatcgaagatgctgcaaga 900
Db 1321 AACGAACAAGAAATTAATCTATCCCTGGAAGAACCTTAAAGATCAACGATGCTGTCAAGA 1380
QY 901 tatgaataaattgggaagaataaacaacgacatacaacatggaagcgtatcttatactgaa 960
Db 1381 TATGAAAATGGGAAAGATATAAACAGCATATCAACACTGGAGCATTTCTTATCTGAA 1440
QY 961 gaaggaagagacatttaaaaaagctgcagatctcctattgagccaaagaagaatgacata 1020
Db 1441 GAAGGAAGAGCATTTTAAAAAACCTTCAGATTCCTATTTGAGCCAAAGAAAGATGACATA 1500
QY 1021 attcattccttatactcaagaagaagaagagctctcaaaagaatacaaatgtatagtagt 1080
Db 1501 ATTCAATCTTTATCTCAAGAAAGAAAGACCTCTTAAAAAGATACAAATGTGATGATGAT 1560
QY 1081 gatttttatactctctgaggaagaagaagcttcttaaaagctacaaatgtatctgtagt 1140
Db 1561 GATTTTATCTCTAGTGAAGAAAGAGTCTTTTAAAAAGCTCAAAATTTGATTCGTGAT 1620
QY 1141 tcttattctgaagaagaagaagacatttaaatagaataaagaatgagtagtagtaactc 1200
Db 1621 TCTTTATCTGAAGAAAGAAAGACCTTTTAAATGAAATACAGGTGATGATGATGATGAT 1680
QY 1201 ttctcgtgaagaagaagaagcttcttaaaagctgaacatgatatctcaacatatgtagt 1260
Db 1681 TTAATCTGAAGAAAGAAAGAGTTTAAAAAGCTGAACCTTGATATTCAACCATATGAT 1740
QY 1261 attaatcaagaagtttgaagaatacagaagaggttaattgatagttcgtaactatctgtagt 1320
Db 1741 ATTAAATCAAGAGTTGCAAGATACAGAGGCTTAATTGATAGTCCGCAATTAATCTTGAT 1800
QY 1321 gtaagaagaagcaglataaagaagatctcaaaatattgtagcttatttcaataccatc 1380
Db 1801 GTAAGAAACAGCATATTAAGAAAGGATATTCAAAATATTGATGCTTTTATTCATCAATTC 1860
QY 1381 ggaagctactctgtacaataaatactatctgtatgtaaaatataatgaataataaacttaca 1440
Db 1861 GGAAGTACCTGTTTACAAATAATTTATTTGATGAAATATGAATATCAATTAACCTTTACA 1920
QY 1441 gcaaccctcagtgtagatattgtagtattccactgttaataacttaataatgaagtagt 1500
Db 1921 GCAACCTTAGTGGCGATTTAGTTGATTCACATGTAATATCTAAATAATTAAGAGCTGAT 1980
QY 1501 ttcaatgaattcaaaataaatactcaaatatagtaattcttagtaactatataatgtagt 1560
Db 1981 TTCAATGAATCAAAAAAATTTCAAAATATGATTTTCTAATACCTATATGATTTGAT 2040
QY 1561 ataaatgaagaagcctgcataagataatgagcgttctgaagaatgagaatccaaatctacca 1620
Db 2041 ATTAATGAAAGGCTCGATGATGATATGAGAGCTTTGAATGGAATGGAATTCATTCACCA 2100
QY 1621 gatactcgaagcagatatattagaanaatggaagccttatatacaagaagaacatcgtctg 1680
Db 2101 GATACCTGAGCAGGATATTTAGAAAGATGAAAGCTTATATTTACAAAGAACATCGGCTG 2160

QY 1681 gaaataaaggatgtacaataaataaagcaatccgaaagaaatataaagatgtgcg 1740
|||||
Db 2161 GAAATTAAGATGTACAAATTAATTAAGCAATCGAAAGAAATATTAAGATGTGATCG 2220
QY 1741 aaagtgtgccaaagataaataagatacaaaaatccaagaagcacagttaataaat 1800
|||||
Db 2221 AAAGTGTGCCAAAGATAAATAGATACAAAATTCAGAAAGCACAGTTAAATATTAAT 2280
QY 1801 caggaaatgaataaagcattaggtgtaccaaaatatcaagaagcttaattcaacgcg 1860
|||||
Db 2281 CAGGAATGAATTAAGCAATTAAGGTTACCAAAATATACAAAGCTTATACATTCAACGTC 2340
QY 1861 cataatagataatgcatccaatattgagaagtgctatttaataatgaaatgaa 1920
|||||
Db 2341 CATTAATAGATATGCATCCAAATATGTAGAAAGCTTATTAATATGTAATGAAATGAAA 2400
QY 1921 aataatattcaaaagtatcttaaaaaaagtaacaataacttaagtgatgaa 1980
|||||
Db 2401 AATTAATATTCAAAAGTGCATCTTATTAATAAAGTAACTTACTTATGATGTTAATGGA 2460
QY 1981 aaattgttttaacgatatctctcctaataatagctgaacaataatacaactaaat 2040
|||||
Db 2461 AATTTTGTTTTACCGATATTAATCTCTCCCTAATATAGCTGAACAATATACACATCAAGAT 2520
QY 2041 gagatataagcaagttcatccaagaagggtatattgtccagaatcccgcttatata 2100
|||||
Db 2521 GAGATATAGCAACAACTTATTCAAAAGGTTATATGTTCCAAATCCCGTTCTATATTA 2580
QY 2101 ctccatgtgacctcaaaagtgtagaaataaagaatgtagagggctttatatacga 2160
|||||
Db 2581 CTCCTAAGGACCTTCAAAAGGTAGAAATTAAGGAATGATAGCGGTTTATACACGAA 2640
QY 2161 tttagacatgctgtgatatattatgctgatatctatagataagaaccaactgatta 2220
|||||
Db 2641 TTTGGCATGCTGTGATATTAATGCTGATATCTATTAATAGATAGAACCAATGTGATTA 2700
QY 2221 gttaacaattctaaanaattctatcttaaaagaagaaggagtaattcaacttcg 2280
|||||
Db 2701 GTTACAAATTTCTAAATAATTCATGATATTTTAAAGGAAGAGATTTAACTGTCG 2760
QY 2281 tatggagaacaataagcgggaattttctgcagaagccttaagttaatgcatctcag 2340
|||||
Db 2761 TATGGGAGAAACAATGAAACGGAATTTTGTGCGAAGCCTTTATGATGATCTACG 2820
QY 2341 gaccatgctgaagcgtttaaagttcaaaaagtaaaagctcgcgaaaacttccaattataac 2400
|||||
Db 2821 GACCATGCTGAACGTTTAAAGTTCAAAAATAATGTCGGAACCTTCCAAATTTATTAAC 2880
QY 2401 gatcagatcaagttcatataactcataa 2430
|||||
Db 2881 GATCAGATTAACTTCATTTAATCACTATA 2910

RESULT 2
LOCUS 13395 3291 bp DNA PAT 06-FEB-1997
DEFINITION Sequence 1 from patent US 5591631.
ACCESSION 13395
VERSION 13395.1 GI:1824186
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3291)
AUTHORS Leppla,S.H., Klumpel,K.R., Arora,N., Singh,Y. and Nicholls,P.J.
TITLE Anthrax toxin fusion proteins, nucleic acid encoding same
JOURNAL Patent: US 5591631-A 1 07-JAN-1997;
FEATURES
source 1..3291
BASE COUNT 1333 a 391 c 564 g 1003 t
ORIGIN

Query Match 100.0%; Score 2430; DB 6; Length 3291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgaataaaaaaagaattataaagaatataatagatgtcatgtttagtaacagcaat 60
|||||
Db 481 ATCAATATATAAAGAAATTTATTAAGATATAGATGATGATGATGATGATGATGATGATGAT 540
QY 61 acttgatgtgtccgtcttatacccccctgtacagggcgagcggtcatgtgatgta 120
|||||
Db 541 ACTTTAGTGTCCCGCTTATATCCCTTGTGACAGGGGGCGGCTGATGCTGATGTA 600
QY 121 ggtatcacgttaaaagaaagaaagaaataaagaatgagataaagaaagaaagaa 180
|||||
Db 601 GGTATGACGTAAAGAGAAAGAAAGAAATTAAGATGAGATTAAGAAAGATGAAAGAA 660
QY 181 cgaataaaaaacacaggaagacattaaaggaatcatgaaacacattgtaaaatagaa 240
|||||
Db 661 CGAATATAAACAACAGGAAGCATTTAAAGGAATCATGAAACACATTTGTTAAATAGAA 720
QY 241 gtaaaaggaggaaagcgtttaaaagagcagcagaagaaagccttcttgaaagta 300
|||||
Db 721 GTTAAAGGGGAGAGAGCTTTAAAGAAAGAGCGCAAGAAAGCTTGTGAAAGTACCA 780
QY 301 tctgatgttttagagatgtataaagcaattggaagaaagaaatataattgtgtgat 360
|||||
Db 781 TCTGATGTTTAAAGATGTTAAAGCAATTTGAGAAAGATATATATTTGTGATGCTGAT 840
QY 361 attcaaaaacataatcctttagaagcatatcgtgaagataaagaaataaagaagcat 420
|||||
Db 841 ATTAACAACATATATCTTTAGAAAGCATTTCTGAAGATTAAGAAATAAAGAAAGCAT 900
QY 421 tatggagaagaatgcttataataatgaacatataatgtatgtcaaaagaaagatatgaa 480
|||||
Db 901 TATGGGAAAGATCTTTTATTAATGAAACATTAATGATATGCAAAAGAGATATGAACC 960
QY 481 gtaactgttaacatcttcgcgaagattatgtagaataaactgaagaagcagcaagct 540
|||||
Db 961 GTACTGTGTAATCCAAATCTTGGAAGATTATGTGAAAATATCTGAAGAGCATGMACTGT 1020
QY 541 tattatgaataagtagaataatatacaaggagataatttaagtaaaatlaaicaacat 600
|||||
Db 1021 TATTATGAATAGGTAAAGATATTAATCAAGGATATTTTAAGTAAATTAATCAACATAT 1080
QY 601 cagaatattttagatgtataataataccataaataagcatcgtatccagatgaaagat 660
|||||
Db 1081 CAGAAATTTTATGATGATTAATTAATACCATTAATAAATGCAATCTGATTCAGATGCAAGAT 1140
QY 661 ctttatttactataagccttaaggaaacatcccaagacattctttagaattcttggaa 720
|||||
Db 1141 CTTTATTATTAATACATGCTTAAGGAACATCCACAGACTTTCTGTAGATTTCTTGGAA 1200
QY 721 caaaatagcaatgaggtagaagaagtaattgtcgaaagcctttgcatatataatcgagca 780
|||||
Db 1201 CAAATATAGCAATGAGGTACAGAAAGATATTGCGAAAGCTTTGCAATATATGAGGCCA 1260
QY 781 cagcatcgatgattgttataagccttatgacgcggaagccttataataatgataatc 840
|||||
Db 1261 CAGCATCGTGTATGTTTACAGCTTTATGCAACCGAAGCCTTTTATTAATGATGATTAATTT 1320
QY 841 aacgaacaagaataataatcatcatccttggaaagacttaagaatacaacgaatgctgaaga 900
|||||
Db 1321 AACGAACAGCAATTAATATCTATCTTGGAAAGACTTAATAATATACCGATGCTGTCAAGA 1380
QY 901 tatgaaaaaatggygaagaaagataaacaagcactatcaacaacgagagcgaattctatcgaa 960
|||||
Db 1381 TATGAATAAATGGGAAAGATTAATAAAGCACTATACACACGTGAGGCAATTTTATATCGAA 1440
QY 961 gaaggaagagactttaaagaagctgcagattcctcatatgtagccaaagaagaatgacata 1020
|||||
Db 1441 GAAAGGAAGAGACTTTTAAAGAGCTGACAGATCTTATGAGCCAAAGAAAGATGACATTA 1500

```

OY 1021 atcaatcttctatctcaagaagaagaagcttctaaagaalacaaatgtaagta 1080
    |||||||
DB 1501 ATTCAATCTTATCTCAAGAGAGAGAGCTTCAAGAAATACAAATGATAGTAGT 1560
OY 1081 gatttttctactgaggaagaagctttaaagaagctacaaatgtatctcgat 1140
    |||||||
DB 1561 GATTCTTAACTACTGAGGAGAGAGCTTTTAAAGAGCTACAAATGATATCTCGAT 1620
OY 1141 tcttctctgaagaagaagaagctttaaagaalacaaatgtaagtaagta 1200
    |||||||
DB 1621 TCTTTTCTCAAGAGAGAGAGCTTTTAAATAGATACAGTGGATAGTAAATCC 1680
OY 1201 ttaactgaagaagaagaagctttaaagaagctgaactgtatctcaacata 1260
    |||||||
DB 1681 TTATCTGAAGAGAGAGAGAGCTTTTAAAGAGCTGAACCTGATATCAACAT 1740
OY 1261 ataatcaagaagctgcaagaalacagaaggttaattgtaagctgcaactta 1320
    |||||||
DB 1741 ATTAATCAAGAGCTGCAAGATACAGAGGCTTAATTGATAGTCCGCAATTA 1800
OY 1321 gtaagaagaagctataagaaggtatcaaaatgtatgcttatacatcaatc 1380
    |||||||
DB 1801 GTAAAGAGAGCTATPAAAGGATATTCAAATATGATGCTTTTACATTAATCC 1860
OY 1381 ggaagtaacctgtacaataaaattatgtatgaagaalacaaatcaacttca 1440
    |||||||
DB 1861 GGAAGTACCTGTGTAATTAATTTATTTGATGAGAAATATGAATATCAATAC 1920
OY 1441 gcaacctgaagctgaggaattgattgatactcaactgataactaaataga 1500
    |||||||
DB 1921 GCAACCCCTAGCTGGGATTTAGTATCCACTGATTAATTAATTAATTAAG 1980
OY 1501 ttaactgaatcaaaaataaattcaaatatgatacttctgaactatgattgatt 1560
    |||||||
DB 1981 TTCAATTAATTCAAAAAATTTCAAAATATGATTTCTAGTACTATATGAT 2040
OY 1561 ataaatgaagaagctgactatgataagagcttgaagaaggaatccaattaca 1620
    |||||||
DB 2041 ATAAATGAAGGCTGCTGATTAATGATTAAGAGCTTTGAAAGGAGAAATCA 2100
OY 1621 gatactcgagcaggaattatagaagaatggaagcttatactaaagaaga 1680
    |||||||
DB 2101 GATACCTGACAGGATATTAGAAATGGAAGCTTATTAATTAAGAAACAT 2160
OY 1681 gaaataaagatgtacaataaatttaagcaatccgaaagaalataaaga 1740
    |||||||
DB 2161 GAAATTAAGATGTACAAATTAATTAAGCAATCCGAAAGAAATATTAAG 2220
OY 1741 aaagtatgtgcaagaagtaaaatagaacaaatcaagaagcagcttaata 1800
    |||||||
DB 2221 AAAGTAGTCCAAAGAGTAAATAGATNCAAAATTCAGAGAGCAAGTTAA 2280
OY 1801 caggaatgaagaagaagcattgaagcttaacaaatatacaagaatcata 1860
    |||||||
DB 2281 CAGGATGAAGTAAGCAATTAAGGCTTACCAAAATATACAAACCTTATTA 2340
OY 1861 cataatagatatgcaataatgtagaagaagcttataatgtaatgaatga 1920
    |||||||
DB 2341 CATATTAATATGCTCCAAATTTGTAGAAAGTCTTATTAATTAATTAAG 2400
OY 1921 aataatattcaagaatgtatctataaagaagcttaacaaatcttaagt 1980
    |||||||
DB 2401 AATTAATATTCAAAGTATCTTATTAAGAAAGCTTAACAAATTAATTAAG 2460
OY 1981 agatttttttctacgaatattactctccataataagctgaagaatatac 2040
    |||||||
DB 2461 AGATTGCTTTTAAACGATATTAATCTCCCTAATATTAATTAATTAAG 2520
OY 2041 gagaatataagaagaagcttcaagaagctatattgttcaagaatccgttca 2100
    |||||||
DB 2521 GAATATATGAGCAAGTCAATTCAGAAAGGCTTATATGCTCCAGATCCG 2580
    |||||||

```

```

OY 2101 ctccatgaccttcaaaaagctgtaagaatgaagaatagtgagggcttatacagaa 2160
    |||||||
DB 2581 CTCATGAGACTTCAGAAAGGCTGATTAAGCAATGATAGTGAAGGCTTTAT 2640
OY 2161 ttggaagctgctgtgagatctatgctggaalactatagaatgaagaacat 2220
    |||||||
DB 2641 TTTGGACATCTGTGATGATTAATGCTGATATCTATTAATTAAGAACCAAT 2700
OY 2221 gtacaaatcttaaaaatcatctgatttttaagaagaagaaggaatcaact 2280
    |||||||
DB 2701 GTTACAAATTTCTAAAAAATTCATTAATTTTAAAGAGAGAGGATTAAT 2760
OY 2281 tatggagaagaacaatgaagcgaatttttgcagaagccttaagtgatcact 2340
    |||||||
DB 2761 TATGGAGAGAACAAATGAACCGGATTTTTCAGAGAGCTTAAATGATTA 2820
OY 2341 gacatgctgcaagcgtttaaagctcaaaaatgtctcgaaacttccaat 2400
    |||||||
DB 2821 GACCATGCTGAAGCTTTAAAGTTCAGAAAGAGCTCCGAAACCTTCA 2880
OY 2401 gatcagattaagctatcatatcaactcaaa 2430
    |||||||
DB 2881 GATCAGATTAAGTCTATTAATTAATTAATTAATTAATTAATTAATTA 2910

RESULT 3
LOCUS 169373 3291 bp DNA PAT 04-FEB-1998
DEFINITION Sequence 1 from patent US 5677274.
ACCESSION 169373
VERSION 169373.1 GI:2831495
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 3291)
AUTHORS Leppla,S.H., Klimpel,K.R., Arora,N., Singh,Y. and Nichols,P.J.
TITLE Anthrax toxin fusion proteins and related methods
JOURNAL Patent: US 5677274-A 1 14-OCT-1997;
FEATURES
source 1..3291
BASE COUNT 1333 a /organism="unknown" 564 g 1003 t
ORIGIN

Query Match 100.0%; Score 2430; DB 6; Length 3291;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgaatataaaaaagaattataaagaatagatagctgttgaacagcaat 60
    |||||||
DB 481 ATGAATATTAATAAAGATTTTAAAGATTAATGATGCTGTTGTTAATAC 540
OY 61 acttgatgtgtcccgcttctatccctgtacaggggagggcggtcagtgatga 120
    |||||||
DB 541 ACTTGAATGTGTCCTGCTTATCCCTTGTACAGGGGGCGGCGTCAATGATG 600
OY 121 ggtatgcagctgaagaagaagaagaagaatagaatgaagaatgaagaaga 180
    |||||||
DB 601 GGTATGCACGTAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
OY 181 cgaataaaacacagagagagcatttaagaagaatctgaacacatctgaaga 240
    |||||||
DB 661 CGAATTAACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
OY 241 gtaaaaggaggagagagctgttaaaagaagcagcagaagaagctacttgaa 300
    |||||||
DB 721 GTAAAAAGGAGAGAGAGCTTTAAAAAGAGAGAGAGAGAGAGAGAGAGAG 780
OY 301 tctgatttttaagaatgataaagaatcgaagaagaatataatgtgagtgat 360
    |||||||
DB 781 TCTGATGTTTAAAGATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
    |||||||

```

OY 361 attcacaacatalatccttagaagcattatcgaagataagaanaaaataaagacatt 420
|||||
Db 841 ATTACAAAACATATATCTTTAGAGCATATCTGAAGATATAGAAAATAAAGACATT 900
OY 421 tatgggaagatgcttattacatgacaatlalatalycaaaagaagatagaacc 480
901 TATGGGAAGATGCTTATTTACATGAACATTTATGTATATGCAAAAAGAGATATGAACCC 960
OY 481 gaaactgtataccaatcttcggaagaattatgtagaaaataactgaaagcactggaacgt 540
|||||
Db 961 GATACCTGTATCCAACTCTTCGGAAGATATGTAGAAAATATGAAAAGCACACGAAACGTT 1020
OY 541 tattatgaataatgtagaataattatcaaggatattlaagtaaaaattaatcaacat 600
|||||
Db 1021 TATTATGAAATATAGTAAAGTATTTATCAAGAGATTTTAAAGTAAATTTAATCAACATAT 1080
OY 601 cagaataattttagatglatlaaaataaccatlaaaatgcalctgattcagatggaacaagat 660
|||||
Db 1081 CAGAAATTTTATAGATGTATTAATATACATTAATAAATGCAATCTGATTCAGATGACAAGAT 1140
OY 661 ctttatttactcaatcagcttaaggaacatccacagactttctgtagaattcttgaa 720
|||||
Db 1141 CTTTTATTTACTAATCAGCTTAAGGAACATCCACAGACTTTTCTGTAGATTTCTTGAA 1200
OY 721 caaaatagcaatggaatgatacaagaagatattgagaaagctttgcatattatatacagcca 780
|||||
Db 1201 CAATAATAGCAATGAGGTATACAGAGATTTTCCAAAGCTTTTGCAATTTATATCGAGCCA 1260
OY 781 cagcatcgtatgcttattacagcttatacgacggaagcttlaattacatgataaatt 840
1261 CAGCATCGATGATGTTTACAGCTTATGACACCGAGCTTTTAAATTAATACATGATTAATTT 1320
OY 841 aacgacaagaataaatacctatcccttggaagaacttaagaatcaacgagatgctgcaaga 900
|||||
Db 1321 AACGACAAGAATAATATATCTATCTTGGAAGAACTTAAAGATCAAGATGCTGTCAAGA 1380
OY 901 tatgaaaaatgggaaaaaagataaaacagcaatcaacacactgagcgatcttattatgaa 960
|||||
Db 1381 TATGAAAAATGGGAAAAGATTAANAACAGCACTATCAACACGAGCAATCTTTATCTGAA 1440
OY 961 gaaggaagaagactttaaanaagctgcagatctcatttgagccaagaagaatgacata 1020
|||||
Db 1441 GAAGGAAGAGACTTTTAAAAAGCTGCAGATTTCCATTTAGGCCAAGAAAAGATGACAT 1500
OY 1021 attcattcttattcctaagaagaagaagccttcaaaaagaatacaaatgtagtagt 1080
|||||
Db 1501 ATTCAATCTTTATCTCMAAGAGAAAAGAGCTTTAAAAAGAAATACAAATTTGATAGTAGT 1560
OY 1081 gatttttattactcaggaagaagaagctttttaaanaagcttacaattgatalctgtagt 1140
|||||
Db 1561 GATTTTTTATCTACTAGAGAAAAGAGTTTAAAAAAGCTTACAAATTTGATTTCTGAT 1620
OY 1141 tcttattcgaagaagaagaagccttcaaatagaatacagtgagtagtaactcct 1200
|||||
Db 1621 TCTTTATCTGAGAGAGAAAAGAGCTTTTAAATAGAAATACAGTGTGATTAATATCT 1680
OY 1201 tcatctgaaaagaagaagaagctttaaanaagctgaaactgatalcacaacatatg 1260
|||||
Db 1681 TATATCTGAAAAAGAAAAGAGTTTAAAAAAGCTGAAACTGATTTATTCACCATATGAT 1740
OY 1261 attaatcaaaaggttgaagaatacagaaggttataatgtagtgcgccaattatctgtagt 1320
|||||
Db 1741 ATTAACTCAAGGTTGCAAGATACAGAGGGTTAATTGATATGTCGCAATTTAATCTTGAT 1800
OY 1321 gtaagaagaacagataaagaagatattcaaaaatgtagtgccttattacaatcaatcatt 1380
|||||
Db 1801 GTRAGAAAACAGATATTAAGGGATATTCAAAATATTGATGCTTTATTTACATCATTCATT 1860
OY 1381 ggaagtagcttgcataataaatttattgtagaaaataatgaataatcaataacttaca 1440
|||||
Db 1861 GGAAATGACTCTGTACAAATTAATTTATTTGTATGAAATATGAAATATCAATTAACCTTACA 1920

OY 1441 gcaacctagtgcggaattagttgatttccactgataactaataataaagagtagt 1500
|||||
Db 1921 GCAACCCCTAGCTCGGATTTAGTTGANTCCAGATATATCTAAATTTATATAGAGATT 1980
OY 1501 tccaatgaaltcaaaaaaaattccaatataagtagtattcttagtaactataatgtagt 1560
|||||
Db 1981 TTCAAATGAATTCAAAAAAATTTCAAAATATAGTATTTCTGTATCAATATATGATTTGAT 2040
OY 1561 ataaatgaagagcctgcatagtagtaatgagcgtttgaagaatggaatccaattaca 1620
|||||
Db 2041 ATAAATGAAGGCTCGATTAGATTAATGAGCGTTTGAAATGAGAAATCCAAATTTATCCCA 2100
OY 1621 gatactcgaacagagataatttagaanaatggaagccttataatacaagaagaacatcgctg 1680
|||||
Db 2101 GATACCTGACAGATATTTAGAAAATGGAACCTTATATATACAAAAGAACATCGGCTG 2160
OY 1681 gaaataaaggaatgatacaataatlaagcaatccgaaanaagaatataaagatgtagt 1740
|||||
Db 2161 GAATTAAGAGATGTACAAATTAATTAAGCAATCCGAAAAAGAAATTTAATAGATTTGATGCG 2220
OY 1741 aaagttagtgcgaagaagtaaaatagatagacaanaatcaagaagcaggttaaatataat 1800
|||||
Db 2221 AAAGTAGTGCCAAAGAGATTAATAGATACAAAATTTCAAGACACAGCTTAAATATTAAT 2280
OY 1801 caggaatggaataaagcattaggttaccaaaatatacaaaagcttattacatcaacgtg 1860
|||||
Db 2281 CAGGAATGGAATTAAGCATTTAGSGTTACCAAAATATACAAAGCTTATTTACATTCACAGCTG 2340
OY 1861 cataatagatatgcatccaatattgtagaaagtgctatttaalatgaaatgaatgga 1920
|||||
Db 2341 CATTAATAGATATCATTCATTCATTTGTAGAAAGTGCTATTTAATTTATGATTAATGATGA 2400
OY 1921 aataatattcaaaatgtagcttataaaaaaggttaacaaatctactagttgtagttagtga 1980
|||||
Db 2401 AATTAATATTTCAAAGTGATCTTATTAANAAGCTTAACAAATTTACTAGTTGATGATTAATGA 2460
OY 1981 agattgcttttaccgatatctctcccaatagcgtgaacaaatatacaacatcaagat 2040
|||||
Db 2461 AGATTTGTTTTTACCGATATTTACTCTCCATTAATATAGCTTAACAAATTTACATTCAGAT 2520
OY 2041 gatatatatgagcaagttcatlcaaaaaggttatalatgltccagaatccgtttcata 2100
|||||
Db 2521 GAGATATATAGGCAAGTTCATTCAAAAGGCTTATATGTTCCAGATCCGTTCTATATTA 2580
OY 2101 ctccaatgacctccaagaagtgtagaatlaaggaatgtagttagttagttagttagttagt 2160
|||||
Db 2581 CTCATGAGACCTTCAAAAGGCTGTAAGATTAAGAAATGATATGATGAGGTTTTTATACAGAA 2640
OY 2161 tttaggacatgctgtgtagttagttagttagttagttagttagttagttagttagttagt 2220
|||||
Db 2641 TTTGGACATCTGTGATGATTTATGCTGATATCTATTAATTAAGAACCAATCTGATTTTA 2700
OY 2221 gttaacaatttcaaaaatctcatgatalatttlaaggaagaaggttaatttcaactc 2280
|||||
Db 2701 GTTACAAATCTTAAAAAATTCATTTGATTAATTTTAAAGAGAAAGAGATTAATTAACCT 2760
OY 2281 tatgggaagaacaatgaagcggaaatttttcgagaagcctttaggttaatgtagttagt 2340
|||||
Db 2761 TATGGGGAACAATATGAAAGGGGAATTTTTCAGAAAGCCCTTTAGTTAATGCAATTTACAG 2820
OY 2341 gaccatgctgaagctttaaagttcaaaaatagctccgaagaacttccaattatatac 2400
|||||
Db 2821 GACCATGCTGAGGTTTAAAGTTCAAAAATAATGCTCCGAAAACCTTCCAAATTTATTAAC 2880
OY 2401 gatcagaatlaagttcatattataactcaaa 2430
|||||
Db 2881 GATCAGATTAAGTTCAATTAATTAACCTATA 2910

RESULT 4
BACLEF BACLEF 3631 bp DNA BCT 26-APR-1993
LOCUS B. anthracis plasmid pX01 lethal factor (lef) gene, complete cds.
DEFINITION

```

ACCESSION      M30210
VERSION        M30210.1 GI:143141
KEYWORDS       lethal factor.
SOURCE         B.anthraxis (strain Sterne) plasmid pX01 DNA, clones pLF7 and
               PBKSPF.
ORGANISM       Bacillus anthracis
               Bacteria, Firmicutes; Bacillus/Clostridium group;
               Bacillus/staphylococcus group; Bacillus; Bacillus cereus group.
REFERENCE      Lowe, J.
AUTHORS        A comparison of Bacillus anthracis sequences
JOURNAL        Unpublished
COMMENT        Draft entry and computer-readable sequence for [1] kindly submitted
               by J.Lowe, 29-NOV-1989.
FEATURES
  source        1..3631
               /organism="Bacillus anthracis"
               /db_xref="taxon:1392"
               585..592
               /note="44 region"
  -35-signal    594..599
               /note="44 region"
  -10-signal    620..625
               /note="44 region"
  mRNA         633..>3631
               /product="lethal factor mRNA"
               /note="ribosome binding site"
               685..783
               /note="lethal factor signal peptide"
               685..3114
               /note="lethal factor precursor"
               /transl_start=1
               /transl_table=1
               /protein_id="AA22569.1"
               /db_xref="GI:143142"
               /translation="MNKKKFTKIVISMSCLVTAITLSPGVPLPVLVAGAGHGVDGMH
               VKKKKKDKENKRRKDEERKNTQEBHKEIMKHIVKIEVKEEAVKKEAAKLEKVED
               VLEMYKATGKRTIVDGTTHKHSLESEDKKIDIGKAKALLHEHYVKAEGEYD
               VLTQSEEDYVNTKALNVYIEGKILSRDILSKINQYKFLVDVLTINAKASDSDG
               ODULTNOLKEHPDPSVEFLQNSNEVOEFAKAPAYITFPOHROVLOLYAPEANT
               MCKPNEOEINLSLELKIDRMISRYKMKKIKOHQOMSDSISEBGRLLKLIQITE
               PKKDIHLSOEKELKRIIDSDPSITBEKELKLIQIDINOSLSEBEKELNIO
               IOVDSNPLSEKKEPKLKLIDIOPYDINORLIDGGLIDSPSINLDVROKRRDIO
               NIDALLHOSIGSTLTKLYLENNMINLITATLADLVSDNTKINRIGIFMKKNE
               KITSISNMIVDINERPALDNERLMRIOLSDPTAGYVLENGKLIORIGIEITQD
               IKQSEKEVIRIDAKVPSKIDTKIOEQLININQEMNKALCLPYTKITITNHRVY
               ASNIVESAILINEMKNOSDLIKKIKVNTYLDNGRREFTDTITPNTAEQYTHODEI
               YEOVHSGLYVEPSRIILHGSKVEELRNDESGTHIEGHAVVDYAGVILDKNSDL
               VTNSKFPDIEKESGNSLTSTGRINEAFEAFLHMSHDHAERLKVKKNAPKTFQF
               INDOI.KFTINS"
  mat_peptide   784..3111
               /note="lethal factor"
               /note="lethal factor"
               1990..2064
               /note="inverted repeat"
               1990..2067
               /note="inverted repeat"
               /note="inverted repeat"
               3334..3529
               /note="transcription terminator"
               3334..3436
               /note="inverted repeat"
               /note="inverted repeat"
               3351..3429
               /note="inverted repeat"
               /note="inverted repeat"
               3358..3524
               /note="cruciform"
  BASE COUNT    1455 a 447 c 602 g 1127 t
  ORIGIN         22.3 kb in the 28.4 kb BamHI B fragment.

Query Match      100.0%; Score 2430; DB 1; Length 3631;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 atgaataaataaagaattataaagtaattgtagtgcattgtagtgcagcaatt 60
|||||

```

```

Db 685 ATGAATATATAAAGAAATTTATATAAGTAATTAGTATGTCATGTTAGTAAACGCAAT 744
Qy 61 actttagtgcgcgcgcgtcttattccccccttgcacagggcgggcggtcattgtagtca 120
Db 745 ACTTTGATGGTCCGCTCTTTATCCCTTTGACAGGGGGGGGGCGCTCATGTGATGTA 804
Qy 121 ggtatgcacgtaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 180
Db 805 GGTATGCACGTAAAGAGAGAAAGAGAAATTAAGATGAGCAATTAAGAAAGATGACAGA 864
Qy 181 cgaataaacaacaggaagaagcatttaagaagaatcagaacacacattgtaaaataga 240
Db 865 CGAAATTAACACAGAGAGAGCAATTAAGAGAAATCATGAATGATGTAATTAAGTA 924
Qy 241 gtaaaaggaggaggagcgtgttaaaagaagcagagaagaactcgtgagaagtacca 300
Db 925 GTAAAGGGAGGAGAGCGTTTAAAAAGAGCAGCAGAAAGCTCTTGAGAAAGTACCA 984
Qy 301 tctgatgttttagagatgataaagcaattggaggaagaatatatatctgtgagtgat 360
Db 985 TCTGATGTTTACAGATGTTATTAAGCAATTTGAGGAAAGATATATTTGTGATGATGAT 1044
Qy 361 attcaaaacataatcctttagaagcattatcgaagatcgaagaagaagaagaagaact 420
Db 1045 ATTCAAAACATATATCTTTAGAGCATATATCTGAGATAGAAATTAAGAAACATTT 1104
Qy 421 tatggaagaagtcgttattatcagacatcattgataatgataaagaagaatgaccc 480
Db 1105 TATGGAAGAGTGCTTTATTAATCATGAACATTTATGCAAAAGAGATATGACACC 1164
Qy 481 gtactgttaaccacatctcgaagaagatcgtatgagaagaactgaagaagcagtc 540
Db 1165 GTACTGTATATCCATCTTGGAAGATTTATGTGAAATTTCTGAAAAGGCACTGAAGTT 1224
Qy 541 tattatgaataatgtaagatattatcaagaagataatttaagaataatcaaccacat 600
Db 1225 TATTATGAATATAGTATGATATATCAAGAGATATTTAGTAAATTTATCAACCATAT 1284
Qy 601 cagaatctttagatgattatataatcattaaataatgcatcgtatcagatggaagaat 660
Db 1285 CAGAAATTTTTATATGATATTTAAATACCATTTAAATGCAATCTGATGAGTGAACAT 1344
Qy 661 ctttatttactaatacagcttaagaacatccacagactcttcttgaatcttgga 720
Db 1345 CTTTATTATTAATACGCTTAAAGCAATCCACAGACTTTTCTGTAGAAATCTTGGA 1404
Qy 721 caaaataagaagaaggtacagaaglatctgcgaagcttgcatalatatacgaacca 780
Db 1405 CAAAATAGCAATGACGTACAGAAATTTGCGAAGCTTTTGCAATATATCGAGCCA 1464
Qy 781 cagcatcgatgattttagaagctttagcccggaagctttaaataatcagataaatt 840
Db 1465 CAGCATCGATGTTTACAGCTTTATGACCGGAGCTTTTAATTTACATGATTAATTT 1524
Qy 841 aacgaacaagaataaataatcattccttgaagaagacttaagaacacagagatgcaaga 900
Db 1525 AACGAACAAGAAATTAATCTATCTTGGAAGACTTTAAAGATCAACGATGCTGCAAGA 1584
Qy 901 tatgaataatgggaagaagataaacaacacatacaacatgagcattcttattcga 960
Db 1585 TATGAATAATGGGAATAATTAACACGATTAACACTGAGGAGTCTTATCTGAA 1644
Qy 961 gaaggaagaagactttaaagaagctgcaagttcctattgagccaaagaagaatgacata 1020
Db 1645 GAGGAAGAAGAGCTTTTAAAAAGCTGAGATCTCTATTTAGCCAAAGAAAGATGACATA 1704
Qy 1021 attaatcttatactcaagaagaagaagaagcttcaaaagaatatacaaatgtagtagt 1080
Db 1705 ATTCAATCTTATCTCAAGAGAAAGAGCTTCTAAAGAAATTAATTAATGATAGTAGT 1764
Qy 1081 gatttttattctactgggaagaagagctttaaagaagctacaaatgtagtagt 1140
Db 1765 GATTTTATCTACTGAGGAAAGAGTTTTAAAAAGCTACAAATTTGATATTCGAT 1824

```


CDS
complement(3389..4024)
/codon_start=-1
/transl_table=11
/product="px01-03"
/protein_id="AAD32307.1"
/db_xref="GI:4894219"
/translation="MKREKGMIGVLCVLVLMILLVPLQAAQSVKVDENKLGNDVS
SFKRDLNNRTKTEKDDSPGSSIIQIKIGIYAPARKVGNIFSIISVIGIYVMAW
ITKNGOMKMAQSGMIFSIISMAKSLSYFLSSVFSFPMNIFTFPLDLFKATILIE
TPPMLAGIRLRIRHESITQPEITRYMANRVITGSCLLCMVLAVLPWLFANM"
complement(4024..4653)
/codon_start=-1
/transl_table=11
/product="px01-04"
/protein_id="AAD32308.1"
/db_xref="GI:4894220"
/translation="MNINISINNVYKLEQYVAREKKQVQLRVSKSSNINAYMIFP
EFSEVKISITELDEYIENLSTSVKYVIDRKLSQLQEKONHSPIMTCEVEDVIDE
TIFHTYVKTPTIAKVFKLVCGLNRLDSLKEIHITVLNNINRIEIEFVEMENET
VAVEKEINAMIAIGAYTKANITYAADLYAAEGNEIKLSIKPKLAGV"
complement(4660..4665)
RBS
complement(4660..4665)
CDS
complement(4660..4665)
/codon_start=-1
/transl_table=11
/product="px01-05"
/protein_id="AAD32309.1"
/db_xref="GI:4894221"
/translation="HELLEYEDMYPOQEEVSIPTPEPEPVQOQVADVQKDMDFVMT
TIVLEQEOGLTRVINNSNNKRYMHIHGEIILPENOLLVYMKLEVDSPSYTLVEMSNRO
LGN"
complement(5318..5323)
RBS
complement(5318..5323)
CDS
complement(5318..5323)
/codon_start=-1
/transl_table=11
/product="px01-06"
/protein_id="AAD32310.1"
/db_xref="GI:4894222"
/translation="MIGGDIQLOTLIIONIKSGNOLKSGTEEMTKAASESKRYLANI
GNVITPVPESITQFTNMVYKIPGTHDAFMKDAQVQKAVECTIKEDVPGVAKQV
LEVDEIEDIIRALTTLTNDKRMVLQOKDITWDVYIVEDSKKRGKQKNIIDQY
EPDLFG"
complement(6231..6236)
RBS
complement(6231..6236)
CDS
complement(6231..6236)
/note="reverse transcriptase, IS629 (574 aa) E. coli
(AB011549): 306/543 positive aa (56%)"
/codon_start=-1
/transl_table=11
/product="px01-07"
/protein_id="AAD32311.1"
/db_xref="GI:4894223"
/translation="MNTVMKSAVSPQVOTQMSIQMKEIENTYVKLRQRIYAEQPN
KRVRKLOLRMLRSKANLLSLIKRVQINQGRKTAGIDITNTPEDRYKLHLKGY
SVRNIKAEPVRAVYIPKKNKGRPLGIPVIKDRIFQNMVKNALPEQMEGRFSPMSYGF
REKSAHAPAMANLFLKLSKGTNRAMIFEGDFQCFDNLNHEILSCIEGFPSMNIQ
WLNACIDIKRTFTYETGTPGGIISPLANTALAGMEKELGVRHFPFRDAMALPD
SIGIVRADDPYVONSKREASWAKLOPYLDKRGKLAKEKTRVYHITLDFDLGE
NPROYPTKEGSLFTKPSQSVKAKKESISELFSKHSRISQGLIRKLPLVTGLANY
WSPVAKVIYGDIDSYVHKVMHLKYNHRGAKMINKKYHPDHTGVSDKWLITD
PDNHNOLIRMRMTPDIVRVHLIKRNSPDANDIKRYFAKDEKIFNREPTNSKRILAK
KTKYKRYONSLVGEELPESNHIYKRVKVGKDEYDNLELLHSCSKHMHALLEWGN
GKQLPKVOAYLSOKIPIINSKAVRMTLTFKKFY"
complement(8362..8367)
RBS
complement(8362..8367)
CDS
complement(8362..8367)
/codon_start=-1
/transl_table=11
/product="px01-08"
/protein_id="AAD32312.1"
/db_xref="GI:4894224"
/translation="WLOETIKQDITDQDOTPAIIQLEFGTASHSAFCFAMATINHHREI
QPAVAGANDSTLKSFRRAISWGTSLFEGEGREDSHTYLDKRVANVNSGQEPINTQ
TINKKAIASKEOLETKYVIAIDAEAPADPPYGLPILEWMAKPIYEEMLT
KNLQPLEVYFDKNEFSSISIAQVVLKEDCKEPLSEMIKRSKQCFPQGGTEKINKI

CDS
complement(11549..11554)
RBS
complement(11549..11554)
CDS
complement(11549..11554)
/codon_start=-1
/transl_table=11
/product="px01-09"
/protein_id="AAD32313.1"
/db_xref="GI:4894225"
/translation="MSVLIGISGLIIVWESNRMFRIATVVICLASQLAGTHLE
VLEQSFVHGVDMLSAEQSLPEMKRFIDF"
complement(11847..11852)
RBS
complement(11847..11852)
CDS
complement(11847..11852)
/codon_start=-1
/transl_table=11
/product="px01-10"
/protein_id="AAD32314.1"
/db_xref="GI:4894226"
/translation="MRLGNDLAAGFPPTPLTEGKHLVRLLOMSEKRAYACFPCCGEG
TLIRSPADETKQAGVOHFHYVGLNANRYGAKKEDLDVYRSPSSPMIASHVFLIF
LIMPYNTLERTENAKSEKMEFNLKRAHLYLDEGGIMVYIYIDRRARDISYTLAKN
YELIGLMRGDENEPEEPOFKOCFTIGKRAATIKNDYFNDRANFENNSDIDPVAKH
VSTLQOMVNHKDMWTIPELRHQNKLIFTSVVDYKDAKYGIGKSEGSLAKRLNRNGY
LLOGETAAERAKMPIASQGLILITGVADLDLGEGLLHAVRGSENVYVHSEBOL
EKTREKVVLOKRRAKFITATPGVWEIV"
complement(12956..12961)
RBS
complement(12956..12961)
CDS
complement(12956..12961)
/codon_start=-1
/transl_table=11
/product="px01-11"
/protein_id="AAD32315.1"
/db_xref="GI:4894227"
/translation="MEDNOKSYAAPVMSSESRARVNMRCFEALILPDGKFOYFKP
HALNKEATQETALITIEGIOEKVGVIIIMRVGQGSFTSGRTIYAEPRKILRLK
DYFDEEYVYEPKK"
complement(13574..13579)
RBS
complement(13574..13579)
CDS
complement(13574..13579)
/codon_start=-1
/transl_table=11
/product="px01-12"
/protein_id="AAD32316.1"
/db_xref="GI:4894228"
/translation="MLKNLILSIFSVSLIFVPIPOQDITMTKIVIVFNGLGISI
PITVARMQDLKOTFLIDPKIKMKKEIKKETTENVGLSLKLOKYOGLKRSHEYIN
RSEBHGILREKRMKDKRNATITPINVAEKLIDIDYSLKLWLELDLHOKIULVSY
VTFACEYLNALIPSLGSSSEKREKVALLSRNIOKMWIVLSAGFEKVEKVEQEEYTE
FPLVQOKNRG"
complement(150402..19002)
CDS
complement(150402..19002)
/note="erythrocyte invasion and possible binding protein;
Rhoipry protein (2401 aa), Plasmodium yoelii (U36927);
520/1194 positive aa (43%)"
/codon_start=-1
/transl_table=11
/product="px01-13"
/protein_id="AAD32317.1"
/db_xref="GI:4894229"
/translation="MAKKEIPDIKYEDIIKAATGNLIIFVGAGISMLGLSPSMKG
YAOELLEDAIEKGVIDSLEYKLEKDEPKKILITQKMLEDEKEIIPKSANLEPFKEK
EYKAIYERLYGMAIYITTNWDECLDVLTTPQSNKYSNMDLSEENKLTQSSSET
KVFIDHTELEAELEKNGVYIHHGYSKKEESLWYVOVLEWYGTISQNSHPEYVFL
DOVFNSSVVLVFMGGLFEFEELIEMVLSKTRKPSHTISHWYFSENSENPQWNLKKY
YSNKGVELIPPNITKKNYDNLPIIDMSKVLGEVSGRDRISKTISIDNMKCDDT

|||||
Db 129542 CTGCATGACGACCTTCAAAAGGTGTAGAAATTAAGAAATGATAGTGGCTTTTATACACGAA 129601
QY 2161 cttagacatgctgtgagatgattatgctgagatcatalagataagaaccaatcgtatca 2220
|||||
Db 129602 TTGTGACATGCTGTGGATGATATATCTGATATCTATGATTAAGAACCAATCTGATTTA 129661
QY 2221 gttaacaattcttaaaaaattcattgattatcttaagaagaagaggtgaattcaactctg 2280
|||||
Db 129662 GTTACAAATTTCTMAAAATTCATTGATATTTTAAAGAAAGAGATATTTTAACTTCG 129721
QY 2281 tatggagagacaatgaagcgaactcttttcgaagcccttagtcaatgcatctagc 2340
|||||
Db 129722 TATGGAGAAACAATAGAACGGAATTTTTCGAAAGCCTTTAGTTAATGCAATTCACG 129781
QY 2341 gaccatgctgaacgtttaaaagttcaaaaaaaatgctcgaagaaacttccaattataac 2400
|||||
Db 129782 GACCAATGCTGAAGCTTTAAAGTTCAAAAAAATGCTCGAAACTTCCAAATTTATTTAAC 129841
QY 2401 gatcagattaagtctatctatcaactcata 2430
|||||
Db 129842 GATCAGATTAACTTCATTATTAATCACTATA 129871

RESULT 6
AX085496 1455 bp DNA PAT 09-MAR-2001
LOCUS AX085496
DEFINITION Sequence 7 from Patent WO0112661.
ACCESSION AX085496
VERSION AX085496.1 GI:13275536
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequence.
1 (bases 1 to 1455)
REFERENCE
AUTHORS Youle, R.J., Liu, X. and Collier, R.J.
TITLE Receptor-mediated uptake of an extracellular bcl-x l? fusion
protein inhibits apoptosis
JOURNAL Patent: WO 0112661-A 7 22-FEB-2001;
Department of Health and Human Services (US) ; PRESIDENT AND
FELLOWS OF HARVARD COLLEGE (US)

FEATURES
source
1. 1455
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="genetic fusion"
1. >1455
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC34031.1"
/db_xref="GI:13275537"
/translation="WGSSHHHHSSGLVPRGSHMAGHDVGMHYKEKKNKDNKR
KDERNKTKOEHLEIMKHIVKIEVKGEEAVKKEAEKLEKVPDVEVMYKAGGKI
YIVGDDITKHSLEALSHEDKKIKDIYKGLAKHAYKAEYKPEVAVIOSSEYVE
NTERALANYEIGKILSRDILSKINOPYKFLDVIYNTIKNSDSDGDLFTNOLKEH
PTDSVEFLKONSNEVQVPAKAFAYIEPHRDVLOLYAPAEFYVMKFNQETINS
MSQSNRELVDPLSYKLSOKGYSMQSFSDVENRTEADBGTESEMETTSALINGNSMH
LADSPAVNGATRAHSSSIDAREYTPMAAKVQAKAREGDEFELEKRAFSDLTSQLHTP
GTATQSFQDVVNELEKRDGVNMGRIYAFSPGALCLVESVDKEMOVLVSRIAMMATYL
NDHLEPMIQENGMDTFVELYGNNAAESRKGOER"

BASE COUNT 487 a 266 c 369 g 333 t
ORIGIN

Query Match 31.6%; Score 767.6; DB 6; Length 1455;
Best Local Similarity 99.5%; Pred. No. 7.7e-100;
Matches 770; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 154 gatgagaataagaagaatgagaacgaaataaaacacaggaagaactttaagaa 213
|||||
Db 118 GATGAGAAATTAAGAAAGATGAAAGACGAAATTAACACAGGAAGCATTTTAAAGGAA 177
QY 214 atcatgaacacatctgtaaaatagaagtaaaagggaagaactctttaaaagaagca 273
|||||
Db 178 ATCATGAACACCTTTTAAATAATAGATGAAGGGGAGAGACCTGTTAAAGAGGCA 237
QY 274 gcaagaagctactctgagaagatcacatctgattctttagagatgataaagcaatgga 333
|||||
Db 238 GCAGAAAGCTACTTACGAAAGATACCATCTGATCTTTAGAGATGATTAAGCAATTTGGA 297
QY 334 ggaagaatataatctgtgagatgagatatacaaaacatalactttagaagcatatct 393
|||||
Db 298 GGAAAGATATATATGTGTGATGATGATTAACAAACATATATCTTTAGAACGATATCT 357
QY 394 gaagataagaanaataaagaacatcttctgggaagaatgcttatcaatgaagcttct 453
|||||
Db 358 GAGGATTAAGAAATAATTAAGACATTTATGGAAAGATGCTTTATTCACATGAACATTTAT 417
QY 454 gtaatacgaagaagaatgagatacgaccgtactgttaatccaatctcggagaagatgta 513
|||||
Db 418 GTATATGCAAAAGAGAGATATGAACCCGTACTTGTATCCAAATCTTCGAAAGATTATGTA 477
QY 514 gaaataactgaaagaagcactggaacgtttattatagaatggtgaagatataccaagga 573
|||||
Db 478 GAAATATCTGAAAGAGGACTGAACGTTTATTAAGAAATAGTAAGATATTAATCAAGGAT 537
QY 574 atttaagaataaataatcaacacatalcagaataattttagatgataaataccataaa 633
|||||
Db 538 ATTTTAAGTAAATTAATCAACCATATCAAGAAATTTTATGATGTATTAATACATTAA 597
QY 634 aatgacatctgattcagaatgagaagatccttatttactaatcagcttaagaagacatcc 693
|||||
Db 598 AATGCATCTGATTCACATGACAGACATCTTTTATTTCAATCACTTAAGAGACATCC 657
QY 694 acagacattctctgtagaattcttggagaacaaatagaagaatgaggtagaagaagatttg 753
|||||
Db 658 ACAGACTTTCTGTAGAAATTTCTGGACAAATAAGCAATGAGGTACAAAGATTTTGGC 717
QY 754 aaagctctgcatatataatcgagccacagcatcgatgattttaaagcttatacgaccg 813
|||||
Db 718 AAAGCTTTTGCATATTAATTCAGACGACACATCGATGATTTTACAGCTTATATGACCG 777
QY 814 gaagctttaaatacatgagataaatttaacgaagaagaataaataatctatccttg 867
|||||
Db 778 GAAGCTTTTAATTAATCAATGATTAATTAAGCAACAGAAATTAATATCCATG 831

RESULT 7
I33397 1368 bp DNA PAT 06-FEB-1997
LOCUS I33397
DEFINITION Sequence 5 from patent US 5591631.
ACCESSION I33397
VERSION I33397.1 GI:1824188
KEYWORDS
SOURCE
ORGANISM
Unknown.
Unclassified.
1 (bases 1 to 1368)
REFERENCE
AUTHORS Leppla, S.H., Klimpel, K.R., Arora, N., Singh, Y. and Nicholls, P.J.
TITLE Anthrax toxin fusion proteins, nucleic acid encoding same
JOURNAL Patent: US 5591631-A 5 07-JAN-1997;
FEATURES
source
1. 1368
/organism="unknown"
Location/Qualifiers

BASE COUNT 411 a 316 c 358 g 283 t
ORIGIN

Query Match 31.4%; Score 762.4; DB 6; Length 1368;
Best Local Similarity 99.2%; Pred. No. 4.2e-99;
Matches 766; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 100 gcgggcggtcattgtagttagtgcacgttaaaagagaagaaataaagatgag 159
|||||
Db 1 GCGGGCGGTCAATGCTGATGTAGTATGCACGTAAAGAAAAGAAAATAAATGATG 60
QY 160 aataagagaaagatgaaagacgaataaaacacaggaagacattaaagaaatcag 219
61 AATAAGAGAAAAGATGAGAACGAATAAACAACAGAAAGCATTTAAAGAAATCATG 120
QY 220 aaacacattgttaaaatagaagtaaaagggaggaagctgttaaaaaagggcagcaga 279
|||||
Db 121 AAACACATTTGTAATAATGAAGTAAAGGAGGAAGCTGTTAAAAAAAGAGCACAGAA 180
QY 280 aagctacttggaagaagtcacatctgtagtctttaagagatcttaagacattggaggaag 339
|||||
Db 181 AAGCTACTTGGAAGAGTACCATCTGATGTTTAGAGATGTAAAGCAATTGGAGGAAG 240
QY 340 atataatctgtgagtggtgatatatacaaacatalatctctagaagcattatcgaagat 399
241 ATATATATTGTGATGATGATATTACAAACATATATCTTTAGAAAGCATTTATCTGAAGAT 300
QY 400 aagaaaaaataaaagacattatgggaaagaatgcttattacatgaaacattatgatat 459
301 AAGAAAAAATAAAGACATTTATGGGAAGATGCTTTATTAACATGAACATTATGTATAT 360
QY 460 gcaaaagaagatatgaaacccgtagtctgtaaccatcttggaagattatgtagaagat 519
361 GCAAAAGAGATATGAAACCGTACTTGTATCCATCTTGGAGATTTATGTAGAAAT 420
QY 520 actgaaaaagcactgaaacgttattatagaataggtaagatatatacaaggaatattca 579
421 ACTGAAAAAGCAGCTGAACGTTTATTTATGCAATAGTAGATATTATCAAGGATTTT 480
QY 580 agtaaaatataacacatatacagaataatttagatgtattaaatcacatnaaaatgca 639
481 AGTAAATTAATCAACCATATACAGAAATTTTAGATGTATTAATACCATTTAAATGCA 540
QY 640 tctgattcaagatgacaagaatcttattactaatcagcttaaggaacatcccaagac 699
541 TCTGATTCAATGACAGCAAGATCTTTTATTTACTATACGCTTAAAGAACATCCACAGAC 600
QY 700 ttctctgtagaatctcttgaaacaaatagaacaaataggtacagaagatattgcgaaagct 759
601 TTTTCTGTAGAAATCTTGGAAACAAATATGCAATGAGTACAAAGAAATTTTGGCAAACT 660
QY 760 ttctgcatatatacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 819
661 TTTGCAATATTAATCGAGCCACAGCATCTGATGTTTACAGCTTTATGACACCGGAAGCT 720
QY 820 ttaataatcatgagtaaaatitaacgaacagaataaataatctatccttggaag 871
721 TTTAATTACATGATTAATTTTAACGAACAGAAATTAATCTACTGCGGACG 772
Db
RESULT 8
169375 1368 bp DNA PAT 04-FEB-1998
LOCUS 169375
DEFINITION Sequence 5 from patent US 5677274.
ACCESSION 169375
VERSION 169375.1 GI:2831497
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1368)
AUTHORS Leppla,S.H., Kimpel,K.R., Arora,N., Singh,Y. and Nichols,P.J.
TITLE Anthrax toxin fusion proteins and related methods
JOURNAL Patent: US 5677274-A 5 14-OCR-1997;
FEATURES
source location/Qualifiers
1..1368
/organism="unknown"
BASE COUNT 411 a 316 c 358 g 283 t
ORIGIN

Query Match 31.4%; Score 762.4; DB 6; Length 1368;
Best Local Similarity 99.2%; Pred. No. 4.2e-99;
Matches 766; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 100 gcgggcggtcattgtagttagtgcacgttaaaagagaagaaataaagatgag 159
1 GCGGGCGGTCAATGCTGATGTAGTATGCACGTAAAGAAAAGAAAATAAATGATG 60
Db
QY 160 aataagagaaagatgaaagacgaataaaacacaggaagacattaaagaaatcag 219
61 AATAAGAGAAAAGATGAGAACGAATAAACAACAGAAAGCATTTAAAGAAATCATG 120
Db
QY 220 aaacacattgttaaaatagaagtaaaagggaggaagctgttaaaaaagggcagcaga 279
121 AAACACATTTGTAATAATGAAGTAAAGGAGGAAGCTGTTAAAAAAAGAGCACAGAA 180
Db
QY 280 aagctacttggaagaagtcacatctgtagtctttaagagatcttaagacattggaggaag 339
181 AAGCTACTTGGAAGAGTACCATCTGATGTTTAGAGATGTAAAGCAATTGGAGGAAG 240
Db
QY 340 atataatctgtgagtggtgatatatacaaacatalatctctagaagcattatcgaagat 399
241 ATATATATTGTGATGATGATATTACAAACATATATCTTTAGAAAGCATTTATCTGAAGAT 300
Db
QY 400 aagaaaaaataaaagacattatgggaaagaatgcttattacatgaaacattatgatat 459
301 AAGAAAAAATAAAGACATTTATGGGAAGATGCTTTATTAACATGAACATTATGTATAT 360
Db
QY 460 gcaaaagaagatatgaaacccgtagtctgtaaccatcttggaagattatgtagaagat 519
361 GCAAAAGAGATATGAAACCGTACTTGTATCCATCTTGGAGATTTATGTAGAAAT 420
Db
QY 520 actgaaaaagcactgaaacgttattatagaataggtaagatatatacaaggaatattca 579
421 ACTGAAAAAGCAGCTGAACGTTTATTTATGCAATAGTAGATATTATCAAGGATTTT 480
Db
QY 580 agtaaaatataacacatatacagaataatttagatgtattaaatcacatnaaaatgca 639
481 AGTAAATTAATCAACCATATACAGAAATTTTAGATGTATTAATACCATTTAAATGCA 540
Db
QY 640 tctgattcaagatgacaagaatcttattactaatcagcttaaggaacatcccaagac 699
541 TCTGATTCAATGACAGCAAGATCTTTTATTTACTATACGCTTAAAGAACATCCACAGAC 600
Db
QY 700 ttctctgtagaatctcttgaaacaaatagaacaaataggtacagaagatattgcgaaagct 759
601 TTTTCTGTAGAAATCTTGGAAACAAATATGCAATGAGTACAAAGAAATTTTGGCAAACT 660
Db
QY 760 ttctgcatatatacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 819
661 TTTGCAATATTAATCGAGCCACAGCATCTGATGTTTACAGCTTTATGACACCGGAAGCT 720
Db
QY 820 ttaataatcatgagtaaaatitaacgaacagaataaataatctatccttggaag 871
721 TTTAATTACATGATTAATTTTAACGAACAGAAATTAATCTACTGCGGACG 772
Db
RESULT 9
133398 1425 bp DNA PAT 06-FEB-1997
LOCUS 133398
DEFINITION Sequence 7 from patent US 5591631.
ACCESSION 133398
VERSION 133398.1 GI:1824189
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1425)
AUTHORS Leppla,S.H., Kimpel,K.R., Arora,N., Singh,Y. and Nichols,P.J.
TITLE Anthrax toxin fusion proteins, nucleic acid encoding same
JOURNAL Patent: US 5591631-A 7 07-JAN-1997;

ACCESSION	133399				
VERSION	133399.1	GI:1824190			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
1 (bases 1 to 1524)					
AUTHORS	Leppla,S.H., Kimmel,K.R., Arora,N., Singh,Y. and Nicholls,P.J				
TITLE	Anthrax toxin fusion proteins, nucleic acid encoding same				
JOURNAL	Patent: US 5591631-A 9 07-JAN-1997;				
FEATURES	Location/Qualifiers				
source	1..1524				
BASE COUNT	436 a	374 c	415 g	299 t	
ORIGIN					

Query Match	31.4%	Score 763.4	DB 6	Length 1524
Best Local Similarity	99.9%	Pred. No. 4,2e-99		
Matches 763; Conservative	0	Mismatches 1	Indels 0	Gaps 0

OY	100	gcgggcgctcgttgctgttgaagtcacgtaaaagaagaagaataaagtgcg	159
Db	1	CGGGGCGCTCAGTGTGATGTRNGCATCCAGTAAACAGAACGAAAAATAAGATGAG	60
OY	160	aataagaagaaaaatgaagaacgaatacaacaaggaaagacattbaaggaatcatg	219
Db	61	AATAAGAGAAAAAGATGAAGAACAATAAATAACACAGSAGAGCAATTAAAGSAATCATG,	1200
OY	220	aaacacattgtaaaatagaagtaaaaagggggaagcgtttaaaaaagsgcacgaa	279
Db	121	AAACACATTGTAAAAAATAGAAAGTAAAAGGGGAGGAGACCTTTAAAAAAGAGCGACGAA	180
OY	280	aagctacttggaaagracccatcgcgatggttttagaagatctatbaagaaatttgaagaa	339
Db	181	AAGCTACTTGGAAGAAATGCCATCTGCATGTTTTAAAGATGATTAAGCAATTTGGAGGAAG	240
OY	340	atatcatactgtgcgttgtatatacaaaaaataatacctttaagaagcattatccgaaat	399
Db	241	ATATATATTTGGATGGTGATATATACAAACATATATCTTAGAAGCATTATTCGAAGAT	300
OY	400	aagaaaaaataaaagacatttatgggaagaatgcttatlacatgaacatbtat	459
Db	301	AAGAAAAAAATAAAAAGCATTATTAAGGAAAAAGATGCTTTATATACATGAACATTATGTATAT	360
OY	460	gcaaaagaagaatatlgaaaccgctactgttaatccaatcttcggaaagtatttgaagaaat	519
Db	361	GCAAAGAGAGATATGAAACCCTGTACTTGTAATCCAATCTTGGAAGATTAATGTAAGAAAT	420
OY	520	actgaagaagcacyaacgcttattacttgaabagtgtaagatatatatcaaggaatatcta	579
Db	421	ACTGAAAAAGCACGACGACGTTTATTATGAAATAGGTAAAGATATTATCAAGGATATTTTA	480
OY	580	agtaaaatlaacacacatalcogaaatltttagatgtlattaatataccatlaaaaaatgca	639
Db	481	AGTAAAAATTATCAACCATATCANAAATTTTAAATSTATTAAATACCATTAATAAATGCA	540
OY	640	cttgatctcagatggaacaagaatcctttatcttacctaatcagcttgaagaaatccccagac	699
Db	541	TCTTATTTACAGTGTGACAAAGATCTTTTATTTATCTAATATGAGTTTAAAGAAATCCCCAGAC	600
OY	700	tcttcctgtagaatctctggaaacaatatagcaatbagttaaagaagratcttgcgaagct	759
Db	601	TTTTTCTGTAGAAATCTTGTGAAACAAAATAGCAATGAGGTACNAGANGATTTTGGCAAAAGCT	660
OY	760	tttgcataatataatcgaacagacagcatcgttgatgtttttaaagctttatgacccggaagct	819
Db	661	TTTTCATATATTATATCGAGCCACACATCGATGATGTTTAAACAGCTTTATGCAACCGGAAGCT	720
OY	820	tttaatactacatgataaatttaagcaacaagaataaactctatc	863
Db	721	TTTTTAATTACATGATGAATAATTTAAAGAAACAAGAAATAAATATCTAAC	764

RESULT	12				
LOCUS	169377				
DEFINITION	Sequence 9 from patent US 5677274.			PAT	04-FEB-1998
ACCESSION	169377				
VERSION	169377.1	GI:2831499			
KEYWORDS	.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 1524)				
TITLE	Lepple,S.H., Klimpel,K.R., Atora,N., Singh,Y. and Nichols,P.J				
JOURNAL	Anthrax toxin fusion proteins and related methods				
FEATURES	Patent: US 5677274-A 9 14-OCT-1997;				
source	Location/Qualifiers				
	1..1524				
	/organism="unknown"				
BASE COUNT	436 a	374 c	415 g	299 t	
ORIGIN					

Query Match	31.4%	Score 762.4;	DB 6;	Length 1524;
Best Local Similarity	99.9%	Pred. No. 4.2e-99;		
Matches 763; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

OY	100	gogggcgctcatgtgatcgtatgtagtataagtaaaagaaggaaaataaagtatgaag	159
Db	1	GCGGGCGGTCATGTGGTGAAGTGTAAGTATGCACCTAAACAGAACGAAGAATAAAGATGAC	60
OY	160	aataagagaaaaagatlgaaagacgaataaacaacaggaagagactttaagyaatatcag	219
Db	61	AATAAGACAAAAAGATGGAAGAAGCAAAATPAAACACAGCAGACGACTTTAAAGAAATCATG	120
OY	220	aaacacactctgtaaaaaatagaagtaaagaagggaagaagctctgtaaaaaagaagcgacgaa	279
Db	121	AAACCATTTGTAAAATATGAAAGTAAACGGAGAGAACTGTTAAAAAAGGCCAGCGAA	180
OY	280	aagctacttgagaaagtcacacctgtagtttttagagatgltataaagcaatttggaggaag	339
Db	181	AAGCTACTTGGAGAAAGTACCCTCATGTGATGTTTTAAGATGTTAAGCAATTGGAGCAAG	240
OY	340	ataatatctgtgsgtgtggtatatacaaaaacatatctttaagaagatratctgaagc	399
Db	241	ATATATATTGTGGATGGTGAATTTACAACAAATATATCTTTAGAACATTTATCTGAAGAT	300
OY	400	aagaaaaaataaaagaaacttatcgggaaagaatgccttatcacatgacatatglat	459
Db	301	AAGAAAAAATATAAGACATTTTAGGGGAAGATGCTTTATACATGAAACATTATAGTATAT	360
OY	460	gcaaaagaagatatgaaccgccgtactctgtlaatccaatccctcgagaatatlgtagaaat	519
Db	361	GCAAAAGAGGATPANGAACCCGTCCTTGTTATCCAAATCTTCGGAAAGTTTGTAGAAAT	420
OY	520	actggaagagcagcabaacgcttataatagaanaagtagaagatatacataaggaatatcta	579
Db	421	ACTGAAAGGACACITGACGCTTTATATGCAAAATAGGTAAAGATATTATCAAGGATATTTTA	480
OY	580	agtaaaatlaatcaaccaatalcagaqaatttttagatglatbaaavaccatluaaatlga	639
Db	481	AGTAAAAATTAATCAACCCTATCACAAAATTTTAAATGATTAATAATPACCATTAAAAATGCA	540
OY	640	tctatctcaagtatgacaagaatccttttatctctaagaacttaagyaagactccaagac	699
Db	541	TCTGATTAAGATGGCAACGATCTTTTATTACTATATCAGCTTAAGAGAACATCCCACAGAC	600
OY	700	ttttctcagaatattcttggaaacaaatagcgaatgaagtlacaagaagiatcttgcgaagc	759
Db	601	TTTTCTGTAGAATTCCTTGACAACAAATATGCAATGAGGACAAAGATATTGGCAAAAGCT	660
OY	760	tttgatataatatcagacacagacacatcgtgtagtttttaagccttatlgcacgggaagc	819
Db	661	TTTGGATTTTATTCGACCCACACATCGATGATGTTTTTACAGCTTTTATGACCGGAAGCT	720

	Oy	820	tttaattacatggataaattcaacgacaagaataaaccttc	863
	Db	721	TTTAAATTACATGCATTAATTAAACGAACAAGAAATAAATCTAAC	764
	RESULT	13		
	BACCYAA			
	LOCUS			
	DEFINITION	BACCYAA	2990 bp	DNA BCT
	ACCESSION	Bacillus anthracis edema factor gene (cya), complete cds.		26-APR-1993
	VERSION	M23179		
	KEYWORDS	M23179.1 GI:142814		
	SOURCE	adenylate cyclase; cya gene; edema factor.		
	ORGANISM	Bacillus anthracis DNA.		
		Bacillus anthracis		
		Bacteria: Firmicutes; Bacillus/Clostridium group;		
	REFERENCE	Bacillus/Staphylococcus group; Bacillus; Bacillus cereus group		
	AUTHORS	1 (bases 1 to 2990)		
	TITLE	Escuyer,V., Duflot,E., Sezer,O., Danchin,A. and Mock,M.		
	JOURNAL	Structural homology between virulence-associated bacterial		
	MEDLINE	adenylate cyclases		
	FEATURES	Gene 71, 293-298 (1988)		
		89138004		
	location/Qualifiers			
	source	1..2990		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:48:20 : Search time 112.1 Seconds

(without alignments)
181.053 Million cell updates/sec

Title: US-09-747-521-2_COPY_9_282

Perfect score: 1404

Sequence: 1 KVISMSCLVATILSGPVFI.....RDVLQVAFAPFNVMKENE 274

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1404	100.0	809	22	AAAB7305
2	1282	91.3	456	15	AAAB60180
3	1282	91.3	472	15	AAAB60181
4	1282	91.3	485	22	AAU00222
5	1282	91.3	508	15	AAAB60182
6	1282	91.3	776	15	AAAB60178
7	427	30.4	800	11	AAAR04236
8	129.5	9.2	980	21	AAAB18294
9	129	9.2	1245	21	AAAB18244
10	125	8.9	558	21	AAAB18273
11	122.5	8.7	508	21	AAAB18187

12	122	8.7	481	20	AAV20047	B. burgdorferi ant
13	122	8.7	497	12	AAV20046	B. burgdorferi ant
14	119	8.5	1427	12	AAAB10534	Human 160kd mediat
15	118.5	8.4	652	18	AAAB18010	Plasmodium falcipa
16	118	8.4	687	19	AAAB1586	Truncated restin p
17	118	8.4	1392	20	AAV06999	Restin protein seq
18	117	8.3	1351	21	AAAB18290	Plasmodium falcipa
19	115.5	8.2	561	19	AAAB63043	Streptococcus uber
20	114	8.1	1135	21	AAV84460	Amino acid sequenc
21	113	8.0	477	10	AAV34601	Chlamydia pneumoni
22	113	8.0	593	14	AAAB4669	Ap Seroctype 7 65kd
23	113	8.0	593	21	AAV51774	A. pleuropneumonia
24	113	8.0	593	21	AAV80378	Drosophila kinesin
25	113	8.0	975	19	AAW72746	Plasmodium falcipa
26	113	8.0	1979	21	AAAB18171	Plasmodium falcipa
27	113	8.0	2440	18	AAW20828	H. pylori cytoplas
28	112.5	8.0	1192	21	AAAB18165	Plasmodium falcipa
29	112	8.0	1093	19	AAW40540	Mutant C-beta prot
30	112	8.0	1558	21	AAAB18324	Plasmodium falcipa
31	112	8.0	1643	21	AAAB45564	Virulence gene pro
32	112	8.0	1786	18	AAW24790	P. falciparum live
33	112	8.0	2110	21	AAAB45528	Virulence gene pro
34	111	7.9	700	16	AAAB5731	B. burgdorferi str
35	111	7.9	716	14	AAAB30730	B. burgdorferi 79
36	111	7.9	975	22	AAAB94042	Human protein sequ
37	111	7.9	1164	17	AAAB5781	Group B streptococ
38	111	7.9	1164	19	AAW40537	Group B streptococ
39	111	7.9	1164	21	AAV84459	Amino acid sequenc
40	110.5	7.9	228	20	AAV20077	B. burgdorferi ant
41	110.5	7.9	259	20	AAV20076	B. burgdorferi ant
42	110	7.8	1164	19	AAW40541	Mutant C-beta prot
43	110	7.8	2485	21	AAAB18172	Plasmodium falcipa
44	109	7.8	265	22	AAAB1187	Human polypeptide
45	109	7.8	307	22	AAAB1714	S. epidermidis ope

ALIGNMENTS

RESULT 1	
AAAB7305	Standard; Protein: 809 AA.
ID	AAAB7305 standard; Protein: 809 AA.
XX	AAAB7305;
AC	29-AUG-2001 (first entry)
DT	
XX	
DE	Wild type B. anthracis lethal factor.
XX	
KW	Lethal factor; LF: immunogen; LF4: protective antigen; PA: DNA vaccine.
KW	humoral; cell-mediated; immune memory response.
XX	
OS	Bacillus anthracis.
XX	
FH	Key
FT	Peptide
FT	Location/Qualifiers
FT	1..33
FT	/label= Signal peptide
FT	/note= "Not given in the specification"
FT	Protein
FT	34..809
FT	/label= LF
FT	42..285
FT	/label= LF4
XX	
PN	WO200145639-A2.
XX	
PD	28-JUN-2001.
XX	
PF	21-DEC-2000; 2000WO-US34912.
XX	
PR	22-DEC-1999; 99US-0171459.
XX	
PA	(OHIS) UNITV OHIO STATE RES FOUND.
PA	(GALL/) GALLOWAY D R.

PA (MATE/) MATECZUN A J.
XX
XX Galloway DR, Mateczun AJ;
XX
XX WPI: 2001-408540/43.
DR N-PSDB; AAC86015.
XX
PT Protecting animal against lethal infection with Bacillus anthracis, by
PT administering wild type or mutated form of Bacillus anthracis lethal
PT factor protein or its fragment or a nucleic acid encoding the mutated
PT protein -
XX
PS Claim 3; Fig 1; 33pp; English.
XX
XX This sequence represents the B. anthracis lethal factor (LF). An
CC immunogenic fragment of LF, LF4, can be used to produce an immune
CC response which protects an animal against lethal infection with
CC Bacillus anthracis. DNA encoding the B. anthracis LF can be used
CC in conjunction with DNA encoding the protective antigen (PA) in a
CC DNA vaccine. Using a DNA vaccine which encodes the mutated LF protein
CC or fragment alone or in combination with a DNA encoding the PA protein
CC or its fragment, both components (humoral and cell-mediated) of the
CC immune system are stimulated, which results in longer term immune
CC memory response. The combined use of a mutated LF and PA gene or their
CC fragments results in a higher level of immune response, as judged by
CC overall serum antibody titers for LF and PA antigens, than the use of
CC either LF or PA genes in separate immunizations.
XX
SQ Sequence 809 AA;

Query Match 100.0%; Score 1404; DB 22; Length 809;
Best Local Similarity 100.0%; Pred. No. 7.3e-103;
Matches 274; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KVISMGLVATITLSPVFPLVVGAGHGDMHVKKEKKNDEKRNKDEENKTOEEH 60
DB 9 kvismglvatitlspvfplvvgagghgdmhvkkekknndenkrkeenkqgeeh 68
OY 61 LKEMRHIVKIEVGEBAVKEAEKLEKVPDVLMEYKAIGKITYVDGDTKHISLE 120
DB 69 lkeimrhivkievgeevkkaekllekvpdvlmeykaigkityvldgdtkhisle 128
OY 121 ALSEDDKKRIDYGRDALLHEHYVYAKEGEYPLVYIOSSDYENFEKALNYYEIGKIL 180
DB 129 alseddkkkldiygdallhehyvakegyepylvyiossedyeentekalnvyeygkll 188
OY 181 SRDILSKINOPYOKFLDVLNTIKNASDSDGQDLFTNOLKEHPPTDFSVFELEQNSNEVOE 240
DB 189 srtilskingpyokfldvlnitknasdsdgdlftnqlkehpptdfsvfeleqnsneve 248
OY 241 VFKAFAFYITEPQHRDVLQLYAPEAFNYMDKENE 274
DB 249 vfakafayitepqrdrvliqyapeafnymdkfne 282

RESULT 2
AAR60180
ID AAR60180 standard; Protein; 456 AA.
XX
XX AAR60180;
XX
DT 04-APR-1995 (first entry)
XX
XX LF(1-254)--TR--PE(401-602) toxin fusion protein.
XX
XX Anthrax: Bacillus anthracis; fusion protein; lethal factor.
XX protective antigen; cell killing; targeting; pathogen;
XX intracellular; HIV; human immunodeficiency virus; toxin;
XX pseudomonas; exotoxin.
XX
XX Bacillus anthracis.
XX
OS Pseudomonas sp.

XX
XX W09418332-A.
XX
XX 18-AUG-1994.
XX
XX 14-FEB-1994; 94WO-US01624.
XX
XX 12-FEB-1993; 93US-0021601.
XX 25-JUN-1993; 93US-0082849.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Arora N, Kimpel K, Leppia SH, Nichols PJ, Singh Y;
XX
XX WPI: 1994-279753/34.
XX
XX N-PSDB; AAR60180.
XX
XX Nucleic acid encoding anthrax toxin fusion protein - useful for
PT targeting toxin to specific cells, eg for killing tumour cells
PT or HIV-infected cells
XX
XX Claim 7; Page 86-87; 124pp; English.
XX
XX This sequence is a fusion protein comprising amino acid residues
CC 1-254 of the anthrax protective antigen binding domain of the
CC native anthrax lethal factor, a two residue linker and residues
CC 401-602 of a Pseudomonas exotoxin A activity inducing domain of a
CC second protein. Such toxin fusion proteins may be useful for the
CC specific killing of tumour cells or the killing of cells infected
CC with intracellular pathogens, especially HIV, depending on their
CC components.
XX
SQ Sequence 456 AA;

Query Match 91.3%; Score 1282; DB 15; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.6e-93;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 AGHGDMHVKKEKKNDEKRNKDEENKTOEEHLKEIMRHIVKIEVGEBAVKEAE 85
DB 1 aghgdmhvkkekknndenkrkdeernktgeehlkemrhivkievgeevkkaeae 60
OY 86 KLEKVPDVLMEYKAIGKITYVDGDTKHISLEALSEKKKRIDYGRDALLHEHYV 145
DB 61 klekvpdvlmeykaigkityvldgdtkhislealsedkkkkldiygdallhehyv 120
OY 146 AKEGYEPVLVYIOSSDYENFEKALNYYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 205
DB 121 akegyepylvyiossedyeentekalnvyeygkilsrdilskingpyokfldvlnitkna 180
OY 206 SDSDDGQDLFTNOLKEHPPTDFSVFELEQNSNEVOEFKAFAFYITEPQHRDVLQLYAPEA 265
DB 181 sdsdgdqdlftnqlkehpptdfsvfeleqnsnevevfkafayitepqrdrvliqyapea 240
OY 266 FNYMDKENE 274
DB 241 fnymdkfne 249

RESULT 3
AAR60181
ID AAR60181 standard; Protein; 472 AA.
XX
XX AAR60181;
XX
DT 04-APR-1995 (first entry)
XX
XX LF(1-254)--TR--PE(398-613) toxin fusion protein.
XX
XX Anthrax: Bacillus anthracis; fusion protein; lethal factor;
XX protective antigen; cell killing; targeting; pathogen;
XX intracellular; HIV; human immunodeficiency virus; toxin;
XX

KV	Pseudomonas; exotoxin.
XX	
OS	Bacillus anthracis.
OS	Pseudomonas sp.
XX	
PN	M09418332-A.
XX	
PD	18- AUG-1994.
XX	
PF	14-FEB-1994; 94MO-USO1624.
XX	
PR	12-FEB-1993; 93US-0021601.
PR	25-JUN-1993; 93US-0082849.
XX	
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.
XX	
PI	Arcora N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;
XX	
DR	WPI: 1994-279753/34.
N-PSDB:	AA070182.
XX	
PT	Nucleic acid encoding anthrax toxin fusion protein - useful for
PT	targeting toxin to specific cells, eg for killing tumour cells
PT	or HIV-infected cells
XX	
PS	Example 2: Page 90-92; 124pp; English.
XX	
CC	This sequence is a fusion protein comprising amino acid residues
CC	1-754 of the anthrax protective antigen binding domain of the native
CC	anthrax lethal factor, a two residue linker and residues 398-613 of
CC	a Pseudomonas exotoxin A activity inducing domain of a second
CC	protein. Such fusion proteins may be useful for the specific
CC	killing of tumour cells or the killing of cells infected with
CC	intracellular pathogens, especially HIV, depending on their
CC	components.
XX	
SQ	Sequence 472 AA:
Query Match	91.3%; Score 1282; DB 15; Length 472:
Best Local Similarity	100.0%; Pred. No. 1.6e-93;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	26 AGGHDVGAMHYKKEKKNDENRKNKTQDEHLKEIMKHIVIEVGEBAVKKEAE 85
Db	4 agghdvgmhvkeknkndenkrkkeernktgeehlkeimkhivievgeevkkaee 63
OY	86 KLELEVPSDVLEMYAIGKTIYVDGITKHTISLEALSDKKRKIDYGKDALHEHYV 145
Db	64 kilekvpsdvlemyaigsktiyvdgdltkhtislealsedkkkkidyykdalhehyv 123
OY	146 AKEGEPELVIVOSSDYEENTEKAALNVYYEIGKILSRDLISKINPYOKFLDVLTIRNA 205
Db	124 akegyepvlivqssedyenteekalnvyveigklisrdlsiskinpqgkflvltirna 183
OY	206 SDSDDQDLFTTNOLEKHPDTSEVFEELEQSNKVQVEFAKAFNYIETPPQRDYQLQTAPPA 265
Db	184 sdsdqgdlftntqlkehptdfsevfleeqnsvgevafakafnyieppqrndvqlqypaea 243
OY	266 FNTYMDKENE 274
Db	244 fnymdkfne 252
RESULT 4	
AAU00222	
ID	AAU00222 standard; Protein: 485 AA.
XX	
AC	AAU00222;
XX	
OT	31-MAY-2001 (first entry)
XX	
Lfn-Bcl-XL apoptosis-modifying fusion protein.	

Human; Lf α -Bcl-XL; apoptosis; cancer; spinal muscular atrophy;
anthrax lethal factor; neoplasm; tumour; hyper-proliferation;
Alzheimer's disease; neurodegenerative disorder; stroke;
transient ischemic neuronal injury; spinal cord injury;
Huntington's disease.

Chimeric - Homo sapiens.
Chimeric - Corynebacterium diptheriae.
Chimeric - Synthetic.

Key	Location/Qualifiers
Region	5..10 /note= "6x histidine tag"
Region	21..276 /note= "Anthrax lethal factor amino acids 1 to 255"
Region	277..485 /note= "Bcl-XL amino acids 1 to 209"

WO200112661-A2.

22-FEB-2001.

15-AUG-2000; 2000WO-US22293.

16-AUG-1999; 99US-0149220.

(HARD) HARVARD COLLEGE.
(USSH) US DEPT HEALTH & HUMAN SERVICES.

Youle RJ, Liu X, Collier RJ;

Pt PI Youle RJ, Liu X, Collier RJ;
xx xx WPI; 2001-218343/22.
DR DR N-PDSB; AAS00250.

Novel fusion protein for modifying apoptosis in target cell and
reducing apoptosis after transient ischemic neuronal injury, has two
domains which targets protein to a cell and modifies apoptotic response
of cell -

Claim 4; Page 64-65; 65pp; English.

The sequence represents the amino acid sequence of Lf α -Bcl-XL apoptosis-modifying fusion protein comprising anthrax lethal factor (Lf) sequence fused to Bcl-XL. The functional apoptosis-modifying fusion protein is capable of binding a target cell and integrating into or crossing a cellular membrane of the target cell. The apoptosis-modifying fusion protein comprises at least two domains: the DTR domain, which targets the fusion protein to the target cell and the Bcl-XL domain, which modifies an apoptotic response of the target cell. The fusion protein is useful for modifying (inhibiting or enhancing) apoptosis in a target cell, such as neuron, lymphocyte, cancer, neoplasm, macrophage, epithelial, stem, tumor or hyper-proliferative cell or an adipocyte. It is also useful for reducing apoptosis in a subject after transient ischemic neuronal injury, especially spinal cord injury. The fusion protein may be used to treat various diseases and injury conditions through inhibition or enhancement of apoptotic cellular response, including neurodegenerative disorders such as Alzheimer's disease, Huntington's disease, spinal muscular atrophy, stroke episodes and unregulated cell growth as in tumors and various cancers. The apoptosis-modifying fusion protein can be delivered effectively throughout the body and targeted to selective tissue and cells.

```

Sequence      485 AA;
Query Match          91.3%; Score 1282; DB 22; Length 485;
Best Local Similarity 100.0%; Pred. NO. 1,7e-93;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
AGCHGDVGHHYKERRKNDEKRKDERRKKQOEHLKEIKRIIVIEVKGEEAVKKAEE 85
|||||vvvvvvvvvvvvvkkkndeknkdkrrkkqeehlkelmklivievyygeevykaa 81
22 agghddvgmhvykeknkdendkrkkeerrktqeehlkelmkllvivievygeevykaa 81

```

Qy	86	KLEKVEPDEVLEAMKKAIGKRIYIVDGGITRIHSLSEALSEPKKRIKDIYGDALLHEHYV	145
Db	82	KILKVEPsdvLEmYKaIggKLyIvDgdLtkHSLseALseDKKKIKdYgdaIlheHYV	141
Qy	146	AKEGYEPVLVQSSSEDDYVENTEKALNLYVEIGKILTSRDIISKINOPAKQFLDVLNTIKNA	205
Db	142	akegyepvLIvqsseddyentekALnlyveIgKILtsrdIIsKlnpqyqKfLdvIntIkna	201
Qy	206	SDSGQDILLFTNQLKEHPTDSVFLELQNSNEVQEVAKAFAYITIEQHRDVLQIYAPEA	265
Db	202	sdsdgdqdlIfnqIkehpdtfsveflEqnsnevgeVtakaFayIepqhrdvIqIyapea	261
Qy	266	FNMYMDKENE 274	
Db	262	fnymdkfine 270	
RESULT 5			
ID	AAR60182		
	AAR60182 standard; Protein; 508 AA.		
XX			
AC	AAR60182;		
XX			
DT	04-APR-1995 (first entry)		
XX			
DE	LF(1-254)--TR--PE(362-613) toxin fusion protein.		
XX			
KW	Anthrax; Bacillus anthracis; fusion protein; lethal factor;		
KW	protective antigen; cell killing; targeting; targeting; pathogen;		
KW	intracellular; HIV; human immunodeficiency virus; toxin;		
KW	Pseudomonas; exotoxin.		
XX			
OS	Bacillus anthracis.		
OS	Pseudomonas sp.		
XX			
PN	MO9418332-A.		
XX			
PD	18-AUG-1994.		
XX			
PE	14-FEB-1994; 94MO-US01624.		
XX			
PR	12-FEB-1993; 93US-0021601.		
PR	25-JUN-1993; 93US-0082849.		
XX			
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.		
XX			
PI	Atora N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;		
XX			
DR	WPI; 1994-279753/34.		
NR	N-PSDB; AAO70183.		
XX			
PT	Nucleic acid encoding anthrax toxin fusion protein - useful for		
PT	targeting toxin to specific cells, eg for killing tumour cells		
PT	or HIV-infected cells		
XX			
PS	Example 1; Page 95-96; 124pp; English.		
XX			
CC	This sequence is a fusion protein comprising amino acid residues		
CC	1-254 of the anthrax protective antigen binding domain of the native		
CC	anthrax lethal factor, a two residue linker and residues 362-613 of		
CC	a Pseudomonas exotoxin A activity inducing domain of a second		
CC	protein. Such fusion proteins may be useful for the specific		
CC	killing of tumour cells or the killing of cells infected with		
CC	intracellular pathogens, especially HIV, depending on their		
CC	components.		
XX			
Sequence	508 AA;		
50			

Query Match	91.3%	Score 1282;	DB 15;	Length 508;
Best Local Similarity	100.0%;	Pred. No. 1,86-93;		
Matches 249; Conservative	0;	Mismatches	0;	Gaps 0;

Qy	26	AGHGDVAMHYKKEKKNKNDKKRRDDEPNKQOEHTKMEIMKHUYKIEVKCEANVKKRAAE	85
Db	1	ag9hgdvamyhvkexkknndemkrrkdeeenkqgeehelhmhykvleevkgeeaakkeaae	60
Qy	86	KLELEVPSDVEEMVKAIGKITYIVDGDITTKHISLEAASEDKKRRKTRDYGKDALDHEHYV	145
Db	61	kileevpsdvlemyaigkityivdgditkhtisleaaseckkkktldygdxaallhehyv	120
Qy	146	AKEGEPEVIAVIOSSDVEENEMKALNVYEGLKILSRDILSKINOPKFLDVNTTKNA	205
Db	121	akegepeviaivgsesdeyenekalanvyeaigkillylstdlslklnqpykflvdvntlckna	180
Qy	206	SDSDQDQDLFTNOLKEHPTDSVFELEONSNEVOEFAKAFAYVIEQOHRVDLQVAPEA	265
Db	181	sdsdsgdqlfltnqlkqhptdsvfeleqnsnevgvafakafayvleqhrvdvldqlyapea	240
Qy	266	FRTYMKPFNE 274	
Db	241	frtymkpfne 249	

[illegible]

Query Match	91.3%;	Score 1282;	DB 15;	Length 776;
Best Local Similarity	100.0%;	Pred. No. 3.1e-93;		

	Matches	249;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	26	AGGHDDVGMHVKKKKKNDKRNKRDEERNKKTDEHLKEMKRIIVKIEVKGEEPAVKKEAAE	85							
Db	1	agghddvgmhvkekknkdenkrkdeernktbeehlkemkhlivlevgeeeavvkkeaae	60							
Qy	86	KLLEKVPSPDVLMEYFAIGKRIYVGADITRIHLSLEALSDKRRKIKRIDYKQDLNHEHY	145							
Db	61	kllekvpdvlemykaigkriyivgditlthlslealsedkrrkikridykdallnheyy	120							
Qy	146	AKEGEIEPIVLVIOSSDDYVENVTEKALNYYEIEKIIISDILSKINOPYQKFLDVLNTIKKA	205							
Db	121	akegeiepiylviossddyventekalnvyeyiekiisrdilskinpqyqflvnlntikna	180							
Qy	206	SDSDQDQLFTNQLKEHPTDFSEVFELEONSNSEVQEFAPAFAYIEIPQRRVQLQIYAPPA	265							
Db	181	sdsdqdlftnqlkehptdfseflegnsnevgvfaafayieipqrtrvdlqiyapaa	240							
Qy	266	FNYMDKENE	274							
Db	241	fnymdkine	249							

CC	Sequence	800 AA;
XX		
CC	former species will also protect against infection by the latter.	
CC	cross-react with the enzyme from B.pertussis, hence a vaccine against the	
CC	enzyme of Bordetella pertussis. Antibodies to B.anthraxis adenyl cyclase	
CC	Given sequence contains several regions of close homology with the Cya	
CC	secretion.	
CC	space), the signal peptide having been cleaved off at the moment of	
CC	signal sequence. The mature protein is secreted into the periplasmic	
CC	In vivo the adenyl cyclase protein is synthesised as a precursor with a	
XX		
XS	Claim 8; 13; 23pp; French.	
PT	Nucleotide sequence encoding adenyl cyclase of Bacillus anthracis	
PT	and derived proteins, useful in protective vaccines,also effective	
PT	against pertussis	
XX		
DR	N-NSDB; Q04123.	
PI	Escuyer V, Duflot E, Mock M, Danchin A;	
PI	(INSP) INSTITUT PASTEUR.	
XX		
XX	25-OCT-1989; 89EP-0402949.	
XX	25-OCT-1989; 88FR-0013952.	
PR		
PA		
PN	EP366550-A.	
PD	02-MAY-1990.	
PD		
PF		
XX		
XX		
OS	Bacillus anthracis.	
OS		
KW	adenyl cyclase; pertussis; protective vaccines; signal sequence	
XX		
DE	Adenyl cyclase gene of Bacillus anthracis.	
DT	12-SEP-1989 (first entry)	
AC	AAR04236;	
ID	AAR04236 standard; protein; 800 AA.	
7		
RESULT		

Query Match	30.4%	Score 427;	DB 11;	Length 800;
Best Local Similarity	35.7%	Pred. No. 1.4e-25;		
Matches 100;	Conservative 54;	Mismatches 100;	Indels 26;	Gaps 6;

```

OY 2 VISMCSLTATILTSRPFRLPLVLQGGAGGNDVMNKKHEKKKKDEKKRDERNNKTOEHL 61
Db 13 i1sfv1lfa1tsa1ena-----mnehtsddlkhnhkthekf 59
OY 62 KEIMKHIVKIEVGSBAVKAEMAEKLEKVPDVLWKAIGKITYIDGDTIKHISLEA 121
Db 60 kdslnlvgfctmetlcklqbtglllklkpkdvlelseygeiyfcdialvheiqd 119
OY 122 LSEDKKTKIKDYGKALLHENHVYKBEDEVLPVLQGGSDVLENVEKRLNLYTEGKILS 188
Db 120 lseeksmmsrgekvvpfasfctvekkctcpbl1-nlkdyplnsegskeyvye1g1s 178
OY 182 RDLISKINOPYOKFLD-VLNTIKMAS-DSDGODILFTNOKUE-----HPTDVSVELEQN 234
Db 179 ly1isk-----dks1dpdflnlkelsdssdsdlfsgkfkexkrlenakidlnfken 233
OY 235 SNEVOGEFAKAPAYIIEQHNDVLDLVAPEAFNYMDKNE 274
Db 234 lte1ghatslasyfapdhrtvlelvaupdamfeynmk 273

```

```

RESULT      8
ID          AAB18294
XX          AAB18294 standard; Protein: 980 AA.
AC          AAB18294;
XX
DT          07-NOV-2000 (first entry)

```

DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:152.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOEF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI: 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 362-365; 577pp; English.

The present invention describes proteins and their fragments (I) encoded by chromosome 2 of the human malarial parasite, *Plasmodium falciparum*. Also described are: (1) nucleotide sequences (II) encoding (I); and (2) vaccines against *P. falciparum* infection comprising (I) or (II). (I) and (II) are useful for the development of vaccines against *P. falciparum* infection. (I) and polyclonal antisera or a monoclonal antibody raised to immunogens comprising the sequences of (I), are useful in the detection of infection with *P. falciparum*. Furthermore, (I) (especially when they are refined or secreted or membrane proteins) can aid the identification of drugs to treat or prevent *P. falciparum* infection, or they can be used to identify drug resistance in *P. falciparum*. Sequencing of the *Plasmodium* chromosome 2 and the subsequent identification of proteins encoded by it will help to expand

[illegible]

```

KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KM antimalarial malaria; protozoacide; infection; insecticide.
XX
XX Plasmodium falciparum.
OS
XX WO200025728-A2.
XX
XX PD 11-MAY-2000.
XX
XX PF 05-NOV-1999; 99WO-US26796.
XX PR 05-NOV-1998; 98US-0107131.
XX
XX PA (HOFF/) HOFFMAN S.
XX PA (CARDU/) CARUCCI D.
XX PA (GARDNER M.) GARDNER M.
XX PA (VENTER J C.) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
FI WPI: 2000-365347/31.
DR
XX PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
PS
PS Disclosure: Page 108-110; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAT70078 to AAT70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
SQ Sequence 508 AA;

```

```
XX AC AAY20047;
XX XX
XX DT 19-JUL-1999 (first entry)
XX XX
XX DE B. burgdorferi antigenic protein, t352.aa.
XX XX
XX KM Antigenic protein; vaccine; Lyme disease; infection; detection.
XX OS
XX OS Borrelia burgdorferi.
XX PN WO9859071-A1.
XX PD 30-DEC-1998.
XX PF 18-JUN-1998; 98WO-US12718.
XX PR 03-SEP-1997; 97US-0057483.
XX PR 20-JUN-1997; 97US-0050359.
XX PR 22-JUL-1997; 97US-0053344.
XX PR 22-JUL-1997; 97US-0053377.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (MEDI-) MEDIMMUNE INC.
XX PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX XX
XX DR WPI; 1999-189980/16.
XX DR N-PSDB; AAX61744.
XX PT New isolated Borrelia burgdorferi nucleic acids - used to develop
XX PT products for the diagnosis, prevention and treatment of diseases
XX PT caused by Borrelia, particularly Lyme disease
XX PS
XX PS Claim 12; Page 172; 275pp; English.
XX CC This sequence represents a Borrelia burgdorferi (Bb) protein of the
XX CC invention, which is suitable for use in a vaccine. The Bb polypeptides
XX CC can be used in vaccines for eliciting protective antibodies to members of
XX CC the Borrelia genus, particularly for the use against Lyme disease in
XX CC humans and animals. They can be used for preventing or attenuating an
XX CC infection caused by a member of the Borrelia genus. The products can also
XX CC be used for detection of members of the Borrelia genus.
XX SQ
XX SQ Sequence 481 AA;

Query Match 8.7%; Score 122; DB 20; Length 481;
Best Local Similarity 21.7%; Pred. No. 0.1;
Matches 51; Conservative 49; Mismatches 81; Indels 54; Gaps 8;

QY 33 GMHYKEKKNKDE--NKRKDEBNKTOEHLKIMKHIVYI-----EVKGEA 78
DB 57 gkdiklpenitrdkklpqkrndendklsvenyenklklnkltkqktsenenkkes 116
QY 79 VKKEEA--EKLEVPSPVLEMYKAIGKITYVGDITKHTSLALSDKKIKIDYKGD 136
DB 117 iekrakyeelltnkikneieiklknkpkcedenyeklnenleee-----td 167
QY 137 ALLEHNVYAKE-----GYEPVLVIQSSDYVENTEKALN--VYVEIGKILSRDILS 186
DB 168 ddfednyendelxtnedny-----psnegllnlnkenlnenekyyalnekkiided 221
QY 187 KINOPYKFLDVLTINAKSDSGODLFTNQLKEHPTDSVFERLEONSNEVOEV 241
DB 222 rinenentlidlqrelrnfkkkdn-----scknlleelenlsisgrl 263

RESULT 13
AAY20046
ID AAY20046 standard; Protein; 497 AA.
XX
XX AC AAY20046;
```

```
XX XX
XX DT 19-JUL-1999 (first entry)
XX XX
XX DE B. burgdorferi antigenic protein, t352.aa.
XX XX
XX KM Antigenic protein; vaccine; Lyme disease; infection; detection.
XX OS
XX OS Borrelia burgdorferi.
XX PN WO9859071-A1.
XX PD 30-DEC-1998.
XX PF 18-JUN-1998; 98WO-US12718.
XX PR 03-SEP-1997; 97US-0057483.
XX PR 20-JUN-1997; 97US-0050359.
XX PR 22-JUL-1997; 97US-0053344.
XX PR 22-JUL-1997; 97US-0053377.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (MEDI-) MEDIMMUNE INC.
XX PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX XX
XX DR WPI; 1999-189980/16.
XX DR N-PSDB; AAX61743.
XX PT New isolated Borrelia burgdorferi nucleic acids - used to develop
XX PT products for the diagnosis, prevention and treatment of diseases
XX PT caused by Borrelia, particularly Lyme disease
XX PS
XX PS Claim 12; Page 172; 275pp; English.
XX CC This sequence represents a Borrelia burgdorferi (Bb) protein of the
XX CC invention, which is suitable for use in a vaccine. The Bb polypeptides
XX CC can be used in vaccines for eliciting protective antibodies to members of
XX CC the Borrelia genus, particularly for the use against Lyme disease in
XX CC humans and animals. They can be used for preventing or attenuating an
XX CC infection caused by a member of the Borrelia genus. The products can also
XX CC be used for detection of members of the Borrelia genus.
XX SQ
XX SQ Sequence 497 AA;

Query Match 8.7%; Score 122; DB 20; Length 497;
Best Local Similarity 21.7%; Pred. No. 0.11;
Matches 51; Conservative 49; Mismatches 81; Indels 54; Gaps 8;

QY 33 GMHYKEKKNKDE--NKRKDEBNKTOEHLKIMKHIVYI-----EVKGEA 78
DB 73 gkdiklpenitrdkklpqkrndendklsvenyenklklnkltkqktsenenkkes 132
QY 79 VKKEEA--EKLEVPSPVLEMYKAIGKITYVGDITKHTSLALSDKKIKIDYKGD 136
DB 133 iekrakyeelltnkikneieiklknkpkcedenyeklnenleee-----td 183
QY 137 ALLEHNVYAKE-----GYEPVLVIQSSDYVENTEKALN--VYVEIGKILSRDILS 186
DB 184 ddfednyendelxtnedny-----psnegllnlnkenlnenekyyalnekkiided 237
QY 187 KINOPYKFLDVLTINAKSDSGODLFTNQLKEHPTDSVFERLEONSNEVOEV 241
DB 238 rinenentlidlqrelrnfkkkdn-----scknlleelenlsisgrl 279

RESULT 14
AAR10534
ID AAR10534 standard; Protein; 1427 AA.
XX
XX AC AAR10534;
XX DT 12-APR-1991 (first entry)
```



```

XX Human 160kd mediator of inflammation protein.
DE Mediator of Imflammation; cytokine; Hodgkin's lymphoma; MRP-160.
KM Homo saplens.
OS
XX EP412050-A.
XX 06-FEB-1991.
PD
XX 26-JUN-1990; 90EP-0810481.
XX 05-JUL-1989; 89GB-0015414.
PR
XX (CIBA ) CIBA GEIGY AG.
PA
XX Odink KG, Tarcsay L, Bruggen J, Wiesendanger W, Cerletti N;
PI Sorz C, Dewolf-Peters C, Delabie J;
XX WPI: 1991-038913/06.
DR N-PSDB; AAQ10378.
XX
PT 160 kD human polypeptide mediator or precursor of inflammation
PT polyclonal or monoclonal antibodies to polypeptide treat and
PT diagnose chronic inflammation and hodgkins' lymphoma
XX
PS Claim 3; Page 32; 47pp; English.
XX
XX The protein is a cytokine used to treat chronic inflammatory
CC conditions. It is prepared by chromatographically purifying an
CC optionally pre-purified cell extract, cell supernatant or cell
CC filtrate of stimulated normal human leucocytes or human embryonic
CC epithelial lung cells. Alternatively, the protein can be produced
CC by microorganisms or continuous mammalian cell lines, transformed
CC with plasmids encoding MRP-160. The invention also covers the
CC polypeptide fragment from amino acids 878-1427 and derivatives of
CC the protein in which the amino and/or hydroxyl functions are
CC glycosylated or acylated and have mol. wt.s of 190 and 140kD,
CC respectively.
XX
SQ Sequence 1427 AA;

```

Query Match 8.5%; Score 119; DB 12; Length 1427;
Best Local Similarity 21.7%; Pred. No. 0.7;
Matches 60; Conservative 51; Mismatches 99; Indels 66; Gaps 13;

```

OY 36 VKEKKNKDNKRR-----DEPRNKTOEHLKEIMKHIVKIEVGGEAVK--- 80
   :| | | : : : : :| | | : : : : :| | | : : : : :| | |
DB 692 lkekensteairskidkeedqhlvemedtlnkigaeik--vkelevlqakcneqtkynd 749
OY 81 -----KEAAEKLE-----KVPSDVLEMYKAIGKRIYVDGDKHISLEALSEDKR-- 127
   :| | | | : : : : :| | | : : : : :| | | : : : : :| | |
DB 750 nftsglkateekllldldalrkassegsemkklrgleaaeqi-khlelekaasskas 808
OY 128 -KINDIVYKDALH-----EHYVYAKGEYEPVLIQSESDYVEMTERALNYYIEIGKTL 180
   : : : : :| | | : : : : :| | | : : : : :| | | : : : : :| | |
DB 809 snrelelgreklhlnlgensevsgvketlekeql-i-kektaaaseeasvqgrsm---- 863
OY 181 SRDLSKINOPYOKFLVDVLTNTIKNASDSGDGLFTNOLKEHPYFSVEFLQNSNEVOE 240
   : : : : :| | | : : : : :| | | : : : : :| | | : : : : :| | |
DB 864 -getvynklhqqeegf-----lmalsd-----leklrenladmeakfreakdereql 908
OY 241 VFKAFAVYIEPQHRDVQLQVLAPEAFN--YMDKFE 274
   :| | : : : : :| | : : : : :| | : : : : :| | : : : : :| | |
DB 909 lkak-----eklendlaeimksgdssqgltkmnd 938

```

RESULT 15
AAW18010
ID AAW18010 standard; Protein: 652 AA.

```

AC AAW18010;
XX
XX 08-DEC-1997 (first entry)
DE Plasmodium falciparum sequestrin.
XX
XX Sequestrin; malaria; vaccine; diagnosis.
XX
XX Plasmodium falciparum clone 3D7.
OS
XX
XX Key
FH Location/Qualifiers
FT 9..18
FT Peptide
FT /note= "Asn/Thr-rich decamer repeat"
FT 19..28
FT Peptide
FT /note= "Asn/Thr-rich decamer repeat"
FT 29..38
FT Peptide
FT /note= "Asn/Thr-rich decamer repeat"
FT 42..180
FT Domain
FT /label= "CD36-binding_domain"
FT /note= "(Claim 24)"
FT
FT Misc-difference
FT 111
FT /note= "encoded by GAA"
FT
FT Misc-difference
FT 119
FT /note= "encoded by GAA"
FT
FT Misc-difference
FT 120
FT /note= "encoded by GAA"
FT
FT Misc-difference
FT 340
FT /note= "encoded by CAA"
FT
FT Peptide
FT 542..547
FT /note= "Asn/Asp-rich hexamer repeat"
FT
FT Peptide
FT 548..553
FT /note= "Asn/Asp-rich hexamer repeat"
FT
FT Peptide
FT 554..559
FT /note= "Asn/Asp-rich hexamer repeat"
FT
FT Peptide
FT 560..565
FT /note= "Asn/Asp-rich hexamer repeat"
FT
FT Peptide
FT 566..571
FT /note= "Asn/Asp-rich hexamer repeat"
FT
FT Peptide
FT 572..577
FT /note= "Asn/Asp-rich hexamer repeat"
FT
FT Peptide
FT 578..583
FT /note= "Asn/Asp-rich hexamer repeat"
FT
FT Peptide
FT 584..589
FT /note= "Asn/Asp-rich hexamer repeat"
FT
FT Peptide
FT 590..595
FT /note= "Asn/Asp-rich hexamer repeat"
FT
FT Peptide
FT 596..601
FT /note= "Asn/Asp-rich hexamer repeat"
FT
FT Peptide
FT 602..607
FT /note= "Asn/Asp-rich hexamer repeat"
FT
FT Peptide
FT 608..613
FT /note= "Asn/Asp-rich hexamer repeat"
FT
FT Peptide
FT 614..619
FT /note= "Asn/Asp-rich hexamer repeat"
FT
FT Peptide
FT 620..625
FT /note= "Asn/Asp-rich hexamer repeat"

```

WO9719168-A1.
29-MAY-1997.
15-NOV-1996; 96WO-US17265.
17-NOV-1995; 95US-0559896.
(USSA) USARMRC US ARMY MEDICAL RES MATERIAL COMMAND.
Duffy PE, Ockenhouse CF;
WPI: 1997-298107/27.
DR N-PSDB; AAT67161.
XX New DNA coding for sequestrin protein from Plasmodium falciparum -

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:51:02 ; Search time 72.22 Seconds
(without alignments)
289.003 Million cell updates/sec

Title: US-09-747-521-2_COPY_9_282
Perfect score: 1404
Sequence: 1 KVISMSCLVTAITLSCGVFI.....RDVLQLYAPFAFNMYDKFNE 274

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	1404	100.0	809	1 J00032	anthrax toxin leth
2	442	31.5	800	1 J50029	adenylate cyclase
3	442	31.5	800	1 B59106	hypothetical prote
4	135.5	9.7	1408	2 T45039	hypothetical prote
5	129.5	9.2	583	2 A70380	arginine-tRNA lig
6	129.5	9.2	980	2 E71606	hypothetical prote
7	129	9.2	1245	2 D71613	GAF domain protein
8	125.5	8.9	1526	2 A45605	mature-parasite-in
9	125	8.9	558	2 C71609	hypothetical prote
10	124	8.8	649	2 T27232	hypothetical prote
11	123.5	8.8	580	2 F75103	conserved hypotet
12	122.5	8.7	508	2 E71620	hypothetical prote
13	122	8.7	497	2 H70168	hypothetical prote
14	122	8.7	1225	2 A49464	chromosome segrega
15	122	8.7	1624	2 T25592	hypothetical prote
16	121	8.6	2829	2 A42771	reticulocyte-bindi
17	120.5	8.6	1156	2 B70356	chromosome assembl
18	120	8.5	709	1 HHBY90	heat shock protein
19	119	8.5	2269	1 T28677	hypothetical prote
20	118.5	8.4	378	2 T18486	hypothetical prote
21	118	8.4	1005	2 A64465	microtubule-vesicl
22	118	8.4	1392	2 A43336	restin - human
23	118	8.4	1427	2 S22695	microtubule-vesicl
24	118	8.4	2401	2 T28676	restin - human
25	117	8.3	1169	2 A64505	phosphoryl protein -
26	117	8.3	1351	2 C71607	plis homolog - Met
27	116.5	8.3	652	2 T41162	hypothetical prote
28	116	8.3	520	2 F70350	hypothetical prote
29	116	8.3	705	2 S55133	recombination prot
					heat shock protein

30	116	8.3	1087	2 T30330	gelsoiin-related p
31	115.5	8.2	886	2 H69378	conserved hypotet
32	115.5	8.2	962	1 S57342	endopeptidase La h
33	114.5	8.2	284	2 S24401	tropomyosin 2, ske
34	114.5	8.2	1365	2 T45031	hypothetical prote
35	113.5	8.1	3351	2 T13812	lipophorin - fruit
36	113	8.0	435	2 A86492	hypothetical prote
37	113	8.0	435	2 A86492	hypothetical prote
38	113	8.0	593	2 H72129	hypothetical prote
39	113	8.0	593	2 S27483	transferrin-bindin
40	113	8.0	593	2 B44796	transferrin-bindin
41	113	8.0	605	2 S46833	hypothetical prote
42	113	8.0	805	2 E70474	translation initia
43	113	8.0	975	1 A31497	kinesin heavy chai
44	113	8.0	1280	2 G96796	hypothetical prote
45	113	8.0	2231	2 C71622	hypothetical prote
					hypothetical prote

ALIGNMENTS

RESULT 1
J00032
anthrax toxin lethal factor pXOI-107 precursor - Bacillus anthracis virulence plasmid
C:Species: Bacillus anthracis
C:Date: 31-Mar-1990 #sequence_revision 11-Nov-1994 #text_change 11-May-2000
C:Accession: J00032; C59104
R:Bray, T.S.; Robertson, D.L.
Gene 81, 45-54, 1989
A:Title: Nucleotide sequence and analysis of the lethal factor gene (lef) from Bacill
A:Reference number: J00032; MWID:90034185
A:Accession: J00032
A:Molecule type: DNA
A:Residues: 1-809

A:Cross-references: GB:M29081; NID:g143143; PIDN:AAV9226.1; PID:g143144
R:Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koeh
J: Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of pXOI, the large Bacillus anthracis plasmid harb
A:Reference number: A59091; MWID:99445483
A:Accession: C59104
A:Molecule type: DNA
A:Residues: 1-809 <OKI>
A:Cross-references: GB:AF065404; NID:g4894216; PIDN:AAD32411.1; PID:g4894323
A:Experimental source: strein Sterne
A:Note: Similar to Anthrax toxin lethal factor precursor, lef, plasmid pXOI, B. anthr
C:Comment: This lethal factor of Bacillus anthracis is part of the tripartite protein
her they cause anthrax, an infectious and often fatal disease of cattle, sheep, and o
C:Genetics:
A:Gene: lef; pXOI-107
A:Genome: plasmid
C:Superfamily: anthrax toxin lethal factor; lethal factor amino-terminal homology
C:Keywords: toxin
F:1-33/Domain: signal sequence #status predicted <SIG>
F:34-809/Product: anthrax toxin lethal factor #status predicted <MAY>
F:44-295/Domain: lethal factor amino-terminal homology <LFA>

Query Match 100.0%; Score 1404; DB 1; length 809;
Best Local Similarity 100.0%; Pred. No. 3.2e-72;
Matches 274; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KVISMSCLVTAITLSCGVFIPLVQAGCHDVGHNHVEKKNKDNKRKDEERNKQDEH 60
DB 9 KVISMSCLVTAITLSCGVFIPLVQAGCHDVGHNHVEKKNKDNKRKDEERNKQDEH 68
QY 61 LKTIKMTIVIEVKGEEAVKKEAEKLEKVPDYLEMYKAIGKITYIVDDITKHSLE 120
DB 69 LKTIKMTIVIEVKGEEAVKKEAEKLEKVPDYLEMYKAIGKITYIVDDITKHSLE 128
QY 121 ALSEDDKKIKDIYKDALHEHYVYAKEGYEPVLVIQSSDDYVENEKALNYYEIGKIL 180
DB 129 ALSEDDKKIKDIYKDALHEHYVYAKEGYEPVLVIQSSDDYVENEKALNYYEIGKIL 188

OY 38 EKEKNKDEERKRDDEERNKTOEE---HLKEIMKHIVYKIEV-----KGEAVKKEAEKLEK 90
||||:| | | | : : : | : : | | | | | : : |
DB 286 EKEKHEQEKOKKNEERERKEDERHIEKQROAEARAKALEOEKOEKEDREKATOK 345
OY 91 VPSDYLEMYKAIGKIYIVDGDITKHISLEALSEDKKIKIDYGDALLHEH-----142
| : : | : : | : : | : : | : : | : : | : : | : : |
DB 346 EREHLEHRRQAEARAKALEKEKHEHRLQOEAOKDEERADARAKAKAREAKSREDEV 405
OY 143 -----YYAKGCEYEV-----VYQSEDEVYENTKAL 170
| : : | : : | : : | : : | : : | : : | : : | : : |
DB 406 AEKVAIDPKOVNIPQKYVASEAKKIKVNELEKOEKRDKIVSKVRESEAEIEKAVR 465
OY 171 NYVEIG-----KILSRDLISKINPYOKFELVLTNTKNASDSODGDLFTNOLKEHPD 225
| : : | : : | : : | : : | : : | : : | : : | : : |
DB 466 SEHDAVAKIPIKPRNDRIYDEKLSFER---IRSRKIVSHEPNSDYL---OFGEKGR 519
OY 226 FSVEFLQONSNEOVFAKAFAYYI 250
| : : | : : | : : | : : | : : | : : | : : | : : |
DB 520 VSIDFLD--SNILFEFEKSCAYFL 542

RESULT 11

F75103
conserved hypothetical protein PAB0812 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 02-Feb-2001
C:Accession: F75103
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: F75103
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-880 <KAS>
A:Cross-references: GB:AJ248286; GB:AL096836; NID:g5458366; PIDN:CAB50131.1; PID:g545864
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB0812
C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

Query Match 8.8%; Score 123.5; DB 2; Length 880;
Best Local Similarity 23.3%; Pred. No. 9.7;
Matches 70; Conservative 50; Mismatches 94; Indels 87; Gaps 15;

OY 31 DVGMHVKEKKNKDEERKRDDEERNKTO-----EHLKEIMKHIVKIEVKG-EEA 78
| : : | : : | : : | : : | : : | : : | : : | : : |
DB 413 EITTRIGMEQEKNEEMKAIIEELRAKGCPCVGRRELTEHKKELMEBYT-LEIKKIEE 471
OY 79 VKKEAAE-----KLEF-KVPSDYLEMYKAIGKI--YIVDGDITKHISLEAL 122
| : : | : : | : : | : : | : : | : : | : : | : : |
DB 472 LRTTEERKRLVNTAKLEIKREFSVMDIAEQIKELSKLKGFMLELEOEKEREFEGL 531
OY 123 SE-----DKKIKIDYGDALLHEHYAAKE-----GYEPV---153
| : : | : : | : : | : : | : : | : : | : : | : : |
DB 532 NEEPNKLKELLGLERDLKRIKALEGRKRLIEKKVAKAELEENLHROLRELFSVESEL 591
OY 154 -LVIOSSED---YVENTEKALNVYIEGKILSRDLISKINPYOKFELVLTNTKNASDS 208
| : : | : : | : : | : : | : : | : : | : : | : : |
DB 592 NRIQELREFHDKYVA-AKSSSELEKLNKLEKE-KTELDAFELADAVENIEEKE--647
OY 209 DGODLLFTNOLKEHPDSEVFLQON-----SNEVOEFAKAFAYYIEPQHDVLY 261
| : : | : : | : : | : : | : : | : : | : : | : : |
DB 648 -----AKLDLESKFNEEYEEKRRLVYKLEREVSSILYAR-----LEELKKSVEQIK 694
OY 262 A 262
| : : | : : | : : | : : | : : | : : | : : | : : |
DB 695 A 695

hypothetical protein PFB0225C - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: E71620
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
; Petter, M.J.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.
science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; M0ID:99021743
A:Accession: E71620
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-508 <GAR>
A:Cross-references: GB:AE001381; GB:AE001362; NID:g3845124; PIDN:AAC71834.1; PID:g384
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0225C

Query Match 8.7%; Score 122.5; DB 2; Length 508;
Best Local Similarity 28.4%; Pred. No. 5.7;
Matches 48; Conservative 32; Mismatches 42; Indels 47; Gaps 10;

OY 37 KEKEKN-----KDEKRRKDEERNKTOEEHLKEIMKHIVKIEVKGAEVAKAEKLEKYP 92
||||:| | | | : : : | : : | | | | | : : |
DB 376 KEKOKKKKKKNEKKNKKKKKKKKKEKE-KNKKKKKKKKKKKKKKKKKKKKKK--432
OY 93 SDVLEMYKAIGKIYIVDGDITKHISLEALSEDKKIKIDYGDALLHEHYVYAKGCEY 152
| : : | : : | : : | : : | : : | : : | : : | : : |
DB 433 -----EKEKN-----NGDVLKHV-----ENNLQDV-----ELLYE-----EK 459
OY 153 VLVYQSEDEVYENTKALNVYIEGKILSRDLISKINQ--PYOKFELDV 199
| : : | : : | : : | : : | : : | : : | : : | : : |
DB 460 IIVNTRKDEELSTK--NKYSE--KDIYHDLISEYSNLTQYNSFLDYM 503

RESULT 13

H70168
hypothetical protein B00553 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: H70168
R:Fraser, C.M.; Castens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kellavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Yu
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; M0ID:98065943
A:Accession: H70168
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-497 <KLE>
A:Cross-references: GB:AE001157; GB:AE000783; NID:g2688471; PIDN:AAC66928.1; PID:g268
A:Experimental source: strain B31

Query Match 8.7%; Score 122; DB 2; Length 497;
Best Local Similarity 21.7%; Pred. No. 6;
Matches 51; Conservative 49; Mismatches 81; Indels 54; Gaps 8;

OY 33 GMHVKEKKNKDE---KKRDEERNKTOEEHLKEIMKHIVYKIEV-----EVGGEA 78
| : : | : : | : : | : : | : : | : : | : : | : : |
DB 73 GKDLKLPENIRDKRLPQKRDENDLSVIEYENKIKINIKELTKTNOKTSENENKIES 132
OY 79 VKKEAA--EKLLEKVPDYLEMYKAIGKIYIVDGDITKHISLEALSEDKKIKIDYGD 136
| : : | : : | : : | : : | : : | : : | : : | : : |
DB 133 IEKAKKAYELITNKLKNEIYIEIKLKLKIKPKEDEVYENKINENIEE-----TD 183
OY 137 ALLEHYVYAKE-----GYEPVLVYQSEDEVYENTKALN---YVYIEGKILSRDLIS 186
| : : | : : | : : | : : | : : | : : | : : | : : |
DB 184 DDFEDNVEYNDLEXTNEDNY-----PSNKGITLNNKLENEMKYYAALNEKKIDELD 237

RESULT 12

E71620

OY 187 KINOPYOKFLDVLWTIKNASDSODDLFTNOLKEHPDVSVEFLQNSNEVOE 241
 DB 238 RINMENTIIDLQELRNFKKKN-----SDKNLEIEENLSIGRI 279

RESULT 14

A49464

Chromosome segregation protein SMCI - yeast (Saccharomyces cerevisiae)
 N:Alternate names: probable head-rod-tail protein SMCI; protein YFL008W
 C:Species: Saccharomyces cerevisiae
 C>Date: 03-May-1994 #sequence.revision 03-May-1994 #text_change 20-Jun-2000

C/Accession: A49464; S56246; S62305; S41804
 R:Strunniakov, A.V.; Larionov, V.L.; Koshtand, D.
 J. Cell Biol. 123, 1635-1648, 1993

A:Title: SMCI: an essential yeast gene encoding a putative head-rod-tail protein is requ
 A:Reference number: A49464; MUID:94103320

A:Accession: A49464

A:Molecule type: DNA

A:Residues: 1-1225 <STR>

A:Cross-references: GB:L00602; NID:g172620; PIDN:AAA16595.1; PID:g172621

R:Murakami, Y.; Naitou, M.; Hagihara, H.; Shibata, T.; Ozawa, M.; Sasamura, S.I.; Sasam

submitted to the EMBL Data Library, May 1995

A:Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces ce

A:Reference number: S56186

A:Accession: S56246

A:Molecule type: DNA

A:Residues: 1-1225 <MUR>

A:Cross-references: EMBL:DS0617; NID:g83685; PIDN:BA09230.1; PID:g836746; MIPS:YFL008W

R:Murakami, Y.

submitted to the EMBL Data Library, April 1994

A:Reference number: S62302

A:Accession: S62305

A:Molecule type: DNA

A:Residues: 1-1225 <MUR>

A:Cross-references: EMBL:DJ1600; NID:g83684; PIDN:BA06496.1; PID:g836816

C:Genetics:

A:Gene: SGD:SMCI

A:Cross-references: SGD:S0001886; MIPS:YFL008W

A:Map position: 6L

C:Function:

A:Description: probably involved in chromosome segregation

C:Superfamily: chromosome segregation protein SMCI

C:Keywords: cell division; coiled coil; mitosis; nucleus

Query Match 8.7%; Score 122; DB 2; Length 1225;

Best Local Similarity 21.9%; Pred. No. 18;

Matches 62; Conservative 50; Mismatches 89; Indels 82; Gaps 13;

OY 32 VGMIVKKEKKDKDEERKKTQEEHLEIKKHIVKIEVGE-EAVKKEAEKLEK 90
 DB 796 IGFPIKEVENHSGELMRQ-----OSKELOOLQKQILTVENKLOFETDRLSTQRRYEK 848
 OY 91 VPSVULEWYKAIIGKITYIVDGDITKHISLEALSDKKIKDIYK-DALLHEHYVAKG 149
 DB 849 AQKD-LR-----NAQVEKMSLEQDEVAIEKISIESKLEEHKHNHDEL 891
 OY 150 YEPVLV-----IQSESDYVENTEKALNYYEIGKILSRDILSKINOPYOKF-LDVLNTIKN 204
 DB 892 OKKRVTKOSELSNSEDLEDNMSNL-----QVLKRE-RGCIKEDIKFLDERVYALKN 943
 OY 205 A-----SDSDGQDLFTNOL-----KEHPTDSVEFLQ 233
 DB 944 CKISININIPISSETTIDLPISSTDNENAITISINDINYGKLPKRYKENNTDSARKELEQ 1003
 OY 234 NSNEVOEV-----AKAFAYIEPOHR-----DVLQVAP 264
 DB 1004 KIHVEEELNLOPNARALERYDEAGREYVNNETDQKAE 1046

RESULT 15

T25592

hypothetical protein C3ZE12.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence.revision 15-Oct-1999 #text_change 15-Oct-1999
 C/Accession: T25592

R:Willcox, L.

submitted to the EMBL Data Library, November 1996

A:Description: The sequence of C. elegans cosmid C3ZE12.

A:Reference number: Z20055

A:Accession: T25592

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1624 <WIL>

A:Cross-references: EMBL:U80032; PIDN:AB53680.1; GSPDB:GN00019; CESP:C3ZE12.4

A:Experimental source: strain Bristol N2; clone C3ZE12

A:Gene: CESP:C3ZE12.4

A:Map position: 1

A:Introns: 13/2; 123/2; 175/1; 249/3; 325/3; 357/3; 513/3; 570/1; 617/3; 674/3; 694/1

Query Match 8.7%; Score 122; DB 2; Length 1624;

Best Local Similarity 25.6%; Pred. No. 25;

Matches 58; Conservative 35; Mismatches 60; Indels 74; Gaps 12;

OY 37 KEKKNKDKERKDEE-----RNKTQEEHLEIKM-----KHIVK---IEVKGEEA 78
 DB 1053 KKQDEEEKKRKKKEEMKKNKEEEKLKRDIEAKSLAKGKEDKSALRKSALNMKDEDA 1112
 OY 79 VKKEAEKLEKVPDVLWYKAIIGKITYIVDGDITKHISLEALSDKKIKDIYGNAL 138
 DB 1113 KKKVATD---EKVAAQKARRF-----GIYVAMAG-----KIDETEEAKK----- 1149
 OY 139 LHEHYVAKEGEPEVLVQSESDYVENTEKALNYYEIGKILSRDILSKINOPYOKFLD- 197
 DB 1150 -----KE--EKVLYKSSRLPAKNADKPKRYEVKIPV-----INDDFDKWDE 1191
 OY 198 VLNTIKNASDSODDLFTNOLKEHPDVSVEFL-----EONSNEVOE 240
 DB 1192 IRKOMKSGS-----NQLQSAIKDLSKGLISATEAKSREMEE 1228

Search completed: December 2, 2001, 13:51:05
 Job time: 300 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:51:46 ; Search time 44.8 Seconds

(Without alignments)
224.245 Million cell updates/sec

Title: US-09-747-521-2_COPY_9_282
Perfect score: 1404
Sequence: 1 KVISMSCLVTAITLSPVFI.....RDVQLVAPAEAFNYMDKFE 274

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1404	100.0	809	LEF_BACAN	P15917 bacillus an
2	1404	31.5	800	CYAA_BACAN	P40136 bacillus an
3	129.5	9.2	583	SYR_AOUAE	O67068 aquifex aeo
4	122	8.7	1225	SMCI_YEAST	P32908 saccharomyc
5	121	8.6	2869	RBPI_PLAVB	O00798 plasmodium
6	120	8.5	709	HS82_YEAST	P02829 saccharomyc
7	118	8.4	1427	REST_HUMAN	P30622 homo sapien
8	116	8.3	520	RECN_AOUAE	O66834 aquifex aeo
9	116	8.3	704	HS83_YEAST	P15108 saccharomyc
10	115.5	8.2	318	YZ11_AOUAE	O66405 aquifex aeo
11	114.5	8.2	284	TPME_CHICK	P18441 gallus gall
12	113	8.0	605	APM2_YEAST	P38700 saccharomyc
13	113	8.0	805	IF2_AOUAE	O67825 aquifex aeo
14	113	8.0	975	KINH_DROME	P17210 drosophila
15	112	8.0	1365	TPMA_XENLA	O01174 xenopus lae
16	111.5	7.9	248	TPMN_XENLA	P16591 homo sapien
17	111.5	7.9	822	FEER_HUMAN	P12270 homo sapien
18	111	7.9	1164	BAG_STRAC	P27991 streptococ
19	111	7.9	2349	TPR_HUMAN	P12270 homo sapien
20	110	7.8	387	YCBR_BACSU	P40758 caenorhabdi
21	110	7.8	1130	YL17_CAEEL	O11102 caenorhabdi
22	109.5	7.8	548	CEAK_ECOLI	O47502 escherichia
23	109.5	7.8	729	KAR3_YEAST	P17119 saccharomyc
24	109	7.8	373	FLHF_AOUAE	O67266 aquifex aeo
25	109	7.8	1682	MSPI_PLAF3	P19598 plasmodium
26	108.5	7.7	245	TPM5_CHICK	P49439 gallus gall
27	108.5	7.7	281	TPM6_CHICK	P49439 gallus gall
28	108.5	7.7	959	LOMN_HUMAN	P36776 homo sapien
29	107.5	7.7	284	TPM2_CHICK	P04268 gallus gall
30	107.5	7.7	284	TPM3_CHICK	P08942 coturnix co
31	107.5	7.7	284	TPMS_CHICK	P49436 coturnix gall
32	107.5	7.7	284	TPMS_COTJA	P49437 coturnix co
33	107	7.6	284	TPMA_COTJA	P02559 coturnix co

ALIGNMENTS

RESULT 1	LEF_BACAN	STANDARD:	PRT:	809 AA.
AC	P15917;			
DT	01-APR-1990 (Rel. 14, Created)			
DT	01-APR-1990 (Rel. 14, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	LETHAL FACTOR PRECURSOR (EC 3.4.24.-) (LF).			
GN	LEF.			
OS	Bacillus anthracis.			
OG	Plasmid PX01.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1392;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 34-49.			
RX	MEDLINE=90034185; PubMed=2509294;			
RA	Braeg T.S., Robertson D.L.;			
RT	"Nucleotide sequence and analysis of the lethal factor gene (lef)			
RT	from Bacillus anthracis.";			
RL	Gene=81:45-54(1989)..			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RA	Lowe J.;			
RL	Submitted (APR-1990) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RX	MEDLINE=95154669; PubMed=7851740;			
RA	Koch S.K., Schlavo G., Montecucco C.;			
RT	"Zinc content of the Bacillus anthracis lethal factor.";			
RL	(FEMS Microbiol. Lett. 124:343-348(1994).			
CC	-1- FUNCTION:--ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,			
CC	AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE			
CC	DEATH. LEF IS THOUGHT TO BE A LETHAL FACTOR THAT, WHEN ASSOCIATED			
CC	WITH PA, CAUSES DEATH. LEF IS NOT TOXIC BY ITSELF. PA IS THOUGHT TO			
CC	BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC CELLS, THEREBY			
CC	FACILITATING THE INTERNALIZATION OF LEF OR EF.			
CC	-1- SUBUNIT: SECRETED ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT			
CC	PROTEINS. A PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN			
CC	EDERA FACTOR (EF). NONE OF THESE IS TOXIC BY ITSELF.			
CC	-1- SUBCELLULAR LOCATION: SECRETED.			
CC	-1- SIMILARITY: THE PA-BINDING REGION IS FOUND IN BOTH B.ANTHRACIS EF			
CC	AND LEF.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M34 (ZINC			
CC	METALLOPROTEINASE).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M29081; AAA79216.1; -			
DR	EMBL; M30210; AAA22569.1; -			


```

Db 60 KOSINNLVTEFTNETLKDIOOTODLLKPKDVLLEYSELGEIYFTDIDVHEKELD 119
Oy 122 LSEDDKKKIDYIGKDALHENVYAKEGEPVLVIQSSDDYVENTEKALNYYEIGKIIS 181
Db 120 LSEEEKSNMNSNGEVPAPSRFVEKRETPKLI-IRKDYAINSEOSSEYVEYIGKIS 178
Oy 182 RILSKINOPYOKFLD--VLNTIKNAS--DSDGODLLETFNOLKE---HPTDFSVETLEON 234
Db 179 LDIIRK-----DKSIDPRELNLKSLSDSDSSDLFFSKFEKLELNKKSIDINFIKEN 233
Oy 235 SNEVGEVFAKAFAYYIEPOHVDVLOLYAPEAFNYMDKENE 274
Db 234 LTFEQHAFSLAFSYFYAPDHTVLELYAPDMFEYMKLEK 273

RESULT 3
STR_AQOAE STANDARD: PRT; 583 AA.
ID SYR_AQOAE 067068;
AC 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ARGINTL-TRNA SYNTHETASE (EC 6.1.1.19) (ARGININE--TRNA LIGASE) (ARGRS).
GN ARGS OR AC_923.
OS Aquifex aeolicus.
OC Bacteria: Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Gidman D.E., Overbeek R., Sneed M.J., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus".
RL Nature 392:353-358(1998).
CC -1- CATALYTIC ACTIVITY: ATP + L-ARGININE + TRNA(ARG) = AMP +
CC PHOSPHATE + L-ARGINYL-TRNA(ARG).
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE000714; AAC07033.1; -
DR InterPro; IPR001278; trna-synt_1d.
DR InterPro; IPR001412; trna-synt_1.
DR Pfam; PF00750; trna-synt_1d; 1.
DR PRINTS; PR01038; TRNASYNTHARG.
DR PROSITE; PS00178; AA_TRNA_LIGASE_L; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 123 "HIGH" REGION.
FT SITE 406 "KMSKS" REGION.
FT BINDING 409 ATP (BY SIMILARITY).
SO SEQUENCE 583 AA; 68318 MW; 350807CA9C1220B CRC64;

Query Match 9.2%; Score 129.5; DB 1; Length 583;
Best Local Similarity 21.1%; Pred. No. 1.5; Mismatches 108; Indels 89; Gaps 12;
Matches 68; Conservative 58;

Oy 38 EKKKKKKKKRD---EERNKTOEHLKEIMKHIVKIE---VKGGAAYKKAACKLE 89
Db 178 EKCPERDETFEIKEIFEKDGVRGEYVKEIAERLKLIVGESICRPEFANLKEVREKIIL 237

```

```

Oy 90 K-----VPSVLEMYKAIGKIIYVDGDIKHSLEALSPD-----KKIKD--- 131
Db 238 EESIELYTKKIEPKKPVVDILSNYG--LDLMKEIKEDISLMDISFDWFESESLDSGE 295
Oy 132 -----YKGDALL-----HEHYVAAKGG 149
Db 296 VERILMLLKEKGVYVKKDGLMLKTSFGDDKORVYKRSDDGTYPFASDIAYHNNFKRG 355
Oy 150 YEPVLVIQSS--DYVENTEKALNYYEIGKIISRDILSKINOPYOKFLDY-----L 199
Db 356 FEKVIIVWGDHNGYIPRKAAALKM-LEIPEDMLELILVQMYLFRGKGVKSKRAGTE 414
Oy 200 NTIKNASDSDGOD--LFTFNOLKEHPTDFSVETLEONSE-----VOEVFAKAFAYYTE 251
Db 415 VTIRELLDEVGKDAVFIFLTGRSDPTLPDVEKAKESSENVYVOTAHARISGIFRE 474

Oy 252 POHRDVLQLYAPEAFNYMDKENE 274
Db 475 FKREYKKDVSVEELINVOHLEE 497

RESULT 4
ID SMC1_YEAST STANDARD: PRT; 1225 AA.
AC P32908;
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CHROMOSOME SEGREGATION PROTEIN SMC1 (DA-BOX PROTEIN SMC1).
GN SMC1 OR CHL10 OR YEL008W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94103320; PubMed=8276886;
RA Struninikov A.V., Lartionov V.L., Koshland D.;
RT "SMC1: an essential yeast gene encoding a putative head-rod-tail
RT protein is required for nuclear division and defines a new ubiquitous
RT protein family.".
RL J. Cell Biol. 123:1635-1648(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shidata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.".
RL Nat. Genet. 10:261-268(1995).
CC -1- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC PART OF A CHROMOSOME CONDENSATION MOTOR.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH SMC2 OR OLIGOMERS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.
CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L00602; AAA16595.1; -
DR EMBL; D50617; BAA09230.1; -
DR EMBL; D31600; BAA06496.1; -
DR PIR; A49464; A49464.

```

DR PIR: S41804, S41804.
 DR SGD: S0001886, SMC1.
 DR InterPro: IPR001687, ATP_GRP_A.
 DR InterPro: IPR003405, SMC_C.
 DR InterPro: IPR003395, SMC_N.
 DR Pfam: PF02463, SMC_C; 1.
 DR Pfam: PF02463, SMC_N; 1.
 KW Mitosis; ATP-binding; Coiled coil; Nuclear protein.
 FT NP_BIND 33 40 ATP (POTENTIAL).
 FT DOMAIN 173 489 COILED COIL (POTENTIAL).
 FT DOMAIN 679 1063 COILED COIL (POTENTIAL).
 FT DOMAIN 1057 1061 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 1137 1164 ALA/ASP-RICH (DA-BOX).
 FT MOTAGEN 173 173 S->L: IN TS MUTANT SMC1-2.
 FT MOTAGEN 458 458 N->D: IN TS MUTANT SMC1-1.
 SQ SEQUENCE 1225 AA, 141279 MW, B504017AA0ECCAC8C CRC64;

Query Match Best Local Similarity 8.7%; Score 122; DB 1; Length 1225;
 Matches 62; Conservative 50; Mismatches 89; Indels 82; Gaps 13;

QY 32 VGHVHEKKNKDNKDKERKKTQDEHLKEIKHIVKVEKGE-EAVKKEAEKLLER 90
 DB 796 IGTITVEYENHSGELMKQ-----OSKEIQLOLOKILITVENKLOFETORLSTTORRYEK 848
 QY 91 VPSDVELEMYKAIGKIYVDGDIKHSLEALSEDDKKIKIDYIGK-DALLHEHYVAKEG 149
 DB 849 AQKD-LE-----NAQVEKSLSEOEFALEMKIGSTESKEEIKHNLDEL 891
 QY 150 YEPVVL-----IOSSEDEVTERKALNVYIEGKILSRDILSKINQYOKF-LDVLNTIKN 204
 DB 892 QKKFVTKOSELNSSEDLIEDMNSNL-----QVLKRE-RDGIKEIDKEFDLERVALKN 943
 QY 205 A-----SDSGQDLFTNQ-----KEHPTGFSVEFLQ 233
 DB 944 CITSINIPISSETTIDLPISSTQNEAITISNIDINKGLPKYKKNNTSAREKLEQ 1003
 QY 234 NSNEVOEVE-----AKAFAYIEPQHR-----DVLQIYAE 264
 DB 1004 KHEVEEILNELQPNARALERYDEAGREYVYNNEQLKAE 1046

RESULT 5
 RBPL_PLAVB STANDARD: PRT: 2869 AA.
 AC 000798;

DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
 GN RBPL.
 OS Plasmodium vivax (strain Belen).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=31273;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92315338; PubMed=1617731.
 RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
 RT "A reticulocyte-binding protein complex of Plasmodium vivax
 merozoites";
 RT Cell 69:1213-1226(1992).
 RL -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
 CC HUMAN RETICULOCYTE CELLS.
 CC -1- SUBUNIT: HOMODIMER (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@lsb-sib.ch.

CC EMBL: M88097, AAA29743.1; -
 DR HSSP: P36956, 1AM9.
 DR Malaria; Receptor; Signal; Transmembrane.
 KW Malaria; Receptor; Signal; Transmembrane.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
 FT DOMAIN 18 2807 EXTRACELLULAR.
 FT TRANSMEM 2808 2826 POTENTIAL.
 FT DOMAIN 2827 2869 CYTOPLASMIC.
 FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
 SQ SEQUENCE 2869 AA, 330213 MW, B9DBE442205EBCF CRC64;

Query Match Best Local Similarity 8.6%; Score 121; DB 1; Length 2869;
 Matches 68; Conservative 42; Mismatches 128; Indels 78; Gaps 11;

QY 31 DVGHHVKEKKNKDNK-----KDEERKKTQDEHLKEIKHIVK-----YVDDG-ITKHI 71
 DB 746 DIALLEVEKKEVTEENKSTLEMLKDEE---MEELQDAKETFAKLNFSDDKLIDVYT 801
 QY 72 ---EYKGEAVKKEAEKLLERVPSDVELEMYKAIGKI-----YVDDG-ITKHI 117
 DB 802 KMSAEVTNAGIKKEIAQGFENVHKKMFSDAFSTFKFALONSQOITNOESDALEKHK 861
 QY 118 SLEALSEDK---KTKIDYIGKDALHEHYVAKESY---EPVLVIOSSSEDEVTERK 168
 DB 862 QNRSEKEEVEFKNSVEEDLSRETEEOEYTKHKNFRRKGEISAEITNMREVLIKIS 921
 QY 169 ALNVYIEI-----KTIISRLSLKINQPYOKFLDVLNTIKNA--- 205
 DB 922 QLNIVYVIEKIFYSLIDQNEVSTAKALKETIVSDSLRDIIDYETEFKERTSAVENTVST 981
 QY 206 --SDSGQDL-----FTNQLKEHPTGFSVEFLQNSNVOEFAKAFAYIEPQHRVL 258
 DB 982 IQSLSKAIDSLKRLNGSINCKKYNTD--IDLLRSIKTLREYQKEMKRDGCENTT 1039
 QY 259 QLYAPAEFVYMDKENE 274
 DB 1040 ALLKSLRDKMGKINE 1055

RESULT 6
 HS82_YEAST STANDARD: PRT: 709 AA.
 AC P02829;

DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HEAT SHOCK PROTEIN HSP82.
 GN HSP82 OR HSP90 OR YPL240C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OX Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84185722; PubMed=6325446;
 RA Fairclly F.W., Finkelstein D.B.;
 RT "Complete sequence of the heat shock-inducible HSP90 gene of
 RT Saccharomyces cerevisiae";
 RT J Biol. Chem. 259:5745-5751(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RA Pohl T.M.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP ATPASE ACTIVITY.
 RX MEDLINE=93123274; PubMed=8419347;
 RA Nadeau K., Das A., Walsh C.T.;

FT DOMAIN 350 1342 COILED COIL (POTENTIAL).
 FT DOMAIN 1408 1421 CCHC-BOX.
 FT VARSPLIC 457 491 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 1069 1069 D -> E (IN REF. 2).
 SQ SEQUENCE 1427 AA; 160989 MW; 0A4F166DD94254E8 CRC64;

Query Match

Best Local Similarity 8.4%; Score 118; DB 1; Length 1427;
 Matches 60; Conservative 51; Mismatches 99; Indels 66; Gaps 13;

QY 36 VKEKKNKDEKNRKK-----DEBNKTKQEEHLEIMKHIVKIGKEGEAVK---- 80
 DB 692 IREKENSLEAIRSKLDKADQHLVEMEDLNKLOEAERK-VKELEVLAKACNEQTKVID 749
 QY 81 -----KEAAEKLE-----KVPDYLEMYKAIGGIIYVDGDTTKHISLEALSDKKK- 128
 DB 750 NTSQTKATEERKLDLALRKASSSECKSEKMLRQOLEAAEKQI-KHLEIEKNAESSKAS 808
 QY 129 --IKNIYKDALH-----EHYVAKEGEPEVLVIOSESEVVENTEKALNYYEIGKIL 180
 DB 809 SITRELOGRELKTMLQENLSEVSQVKETLEKLOI-LKEKFAEASEBAVSQVRSM---- 863
 QY 181 SRDLISKINOPYOKFLDVLNTIKNASDSGDGLLFTNOLKHPDTSVEFLQNSNEVOE 240
 DB 864 -QETVNMHKEEOP-----NMLSSD-----LEKLRNLADMEKFKERDEREQOL 908
 QY 241 VFAKAFAYVTEPOHVDLYLAPAFN--YMDKENE 274
 DB 909 IKAK-----EKLENDIAELIMKSGDSSQTLTKMND 938

RESULT 8

RECN_AQUAE
 ID RECN_AQUAE STANDARD; PRT; 520 AA.

AC 066834;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE DNA REPAIR PROTEIN RECN (RECOMBINATION PROTEIN N).
 GN RECN OR AQ_561.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.
 OX NCBI_Taxid=63363;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=VES;
 RA MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Snead M.A., Keller M., Anujay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.,
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 aeolicus.";
 RL Nature 392:353-358(1998).

-1- FUNCTION: MAY BE INVOLVED IN RECOMBINATIONAL REPAIR OF DAMAGED

DNA (BY SIMILARITY).

-1- SIMILARITY: BELONGS TO THE RECN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL, AE000695; AAC06789.1;
 DR InterPro: IPR003439; ABC_Transportr.
 DR InterPro: IPR003405; SMC_C.
 DR Pfam: PF02483; SMC_C; 1.
 KW DNA repair; ATP-binding; Complete proteome.
 FT NP_BIND 29 36 ATP (POTENTIAL).
 SQ SEQUENCE 520 AA; 60439 MW; A9708562ACBE901E CRC64;

Query Match
 Best Local Similarity 8.3%; Score 116; DB 1; Length 520;
 Matches 61; Conservative 52; Mismatches 78; Indels 48; Gaps 14;

QY 36 VKEKKNKDEKNRKKDE-----RNKTQ-----EHLKIMKHIVKIGKEGEAVKKAER 86
 DB 132 VKLEKVVNSLKRKEDELEFLRKLKQKNLYEFRRREVEIGISSEYEELKNKAN 191
 QY 87 L--LEKVPDYLE--MYKAIGK--IYVDGDTTKHIS-LEALSDKKRKIDYGDAL 139
 DB 192 LNNLEKVKAKVCESTLYKLLEGNSVYEIGIRKLAVESYS--GKFSLEIKIANL 248
 QY 140 HEHYVAKEGEPEVLVIOSESEVVENTEKALNYYEIGILSRDLISKINOPYOK-FLDV 198
 DB 249 EE-----EYVE--LYNSLKEEMPELSEEVN--EINEKLF--TORLEEKYKKSPEI 295
 QY 199 LNTIKNASDSGDGLLFTNOLKHPDTSVEFLQNSNEVOEFAKAFAYVTEPOHND 256
 DB 296 LKEVE-----EIKEELSNLSVDFKEELREVEKUREYKLAEEVSHD 340

RESULT 9

HS83_YEAST
 ID HS83_YEAST STANDARD; PRT; 704 AA.

AC P15106;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE HEAT SHOCK COGNATE PROTEIN HSC82.
 GN HSC82 OR YMR186W OR YMR010.16.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_Taxid=4932;
 RN [1]
 RP SEQUENCE FROM N.A.

RC MEDLINE=89384620; PubMed=2674684;
 RA Borokovich K.A., Farrell F.W., Finkelstein D.B., Taulien J.,
 RA Lindquist S.;
 RT "hsp82 is an essential protein that is required in higher
 RT concentrations for growth of cells at higher temperatures.";
 RL Mol. Cell. Biol. 9:3919-3930(1989).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Boucherie H., Dujardin G., Kermorgant M., Monribot C., Slonimski P.P.,
 RA Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]

SEQUENCE OF 1-7.

CC STRAIN=S288C;

CC MEDLINE=96093904; PubMed=7483834.

CC Boucherie H., Dujardin G., Kermorgant M., Monribot C., Slonimski P.P.,
 CC Perrot M.;
 CC "Two-dimensional protein map of Saccharomyces cerevisiae:
 CC construction of a gene-protein index.";
 CC Yeast 11:601-613(1995).

-1- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY
 (BY SIMILARITY).

-1- INDUCTION: EXPRESSED CONSTITUTIVELY AT A VERY HIGH LEVEL AND IS
 MODERATELY INDUCED BY HIGH TEMPERATURES.

-1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL, M26044; AA02813.1; .

DR EMBL: 249808; CAA89919.1; -
 DR PIR: A32572; A32572.
 DR HSSP: P02829; 1AH8.
 DR SWISS-2DPAGE: P15108; YEAST.
 DR YEPD; 9860; -.
 DR SGD: S0004798; HSC82.
 DR InterPro: IPR003594; HATPase_C.
 DR InterPro: IPR001404; HSP90.
 DR Pfam: PF02518; HATPase_C; 1.
 DR Pfam: PF00183; HSP90; 1.
 DR PRINTS: PR00775; HEATSHOCK90.
 DR SMART: SM00387; HATPase_C; 1.
 DR PROSITE: PS00298; HSP90; 1.
 DR Chaperone; ATP-binding; Heat shock; Multigene family; Repeat.
 FT INIT_MET 0
 FT DOMAIN 220 258
 FT REPEAT 220 224
 FT REPEAT 225 229
 FT REPEAT 230 234
 FT REPEAT 245 249
 FT CONFLICT 618 618
 FT CONFLICT 620 620
 SQ SEQUENCE 704 AA; 80768 MW; 979B5258BDFD53B6 CRC64;

Query Match 8.3%; Score 116; DB 1; Length 704;
 Best Local Similarity 21.1%; Pred. No. 10;
 Matches 60; Conservative 52; Mismatches 101; Indels 72; Gaps 11;

OY 36 VKEKKNDENKRDENK-----TOEHLKEIMKIHYIEKGEFAVYKEAEKLEKV 91
 DB 218 IPEEKKDEKDEKDEKLEVEDEEKKPKTKKAEVEDELELNK--TKPLMTN 275
 OY 92 PSDVLE---MYKAIIG---KIYIVDGDITKHISLEA-----LSKD 125
 DB 276 PSDITQEEYNAFYKKSISNMDPELVY-----KHFSVEQLFEFRAILFIPKRAPDLEFS 329
 OY 126 KKKIDYIGKALLHEHYIAKEGYEPVY-----IQSSEDIYVENTKALNYEEIG 177
 DB 330 KKKKNNI-----KLYVRVFTIDEAEDLIPWLSFYKGVDSDDLPLNLSREMLQONKIM 384
 OY 178 KILSDILSKI-----NOPYOKFLDVL-NTIKNASDSGDGLFTNOLKEHPTDF 226
 DB 385 KTKRNIYKLLLEAFNEIADSEDFKFSAFKNIKLGVEHTONRALAKILRNSTR 444
 OY 227 SVEFLEQNSNEVOEYFAKAFAYIIEPQHDVLQLYAPFAFYMDK 271
 DB 445 SVDELTSITDVTYTRM-----PEHOKNIYITGESLKAVER 479

RESULT 10
 Y211_AQUAE STANDARD; PRT; 318 AA.
 AC 066405;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN AALL.
 GN AALL.
 OS Aquifex aeolicus.
 OC Plasmid ecel.
 CC Bacteria; Aquificales; Aquificaceae; Aquifex.
 OX NCBI_TaxID=63363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VF5;
 RX MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Shedd M.A., Keller M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 aeolicus.";

RL Nature 392:353-358(1998).
 CC -1- SIMILARITY: STROMG, TO A AEOLICUS AA07 AND AA34.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC
 DR EMBL: AE000667; AAC07957.1; -
 DR Hypothetical protein; Plasmid; Complete proteome.
 KW SEQUENCE 318 AA; 38423 MW; 4C2E5B0050C6AC7C CRC64;
 SQ

Query Match 8.2%; Score 115.5; DB 1; Length 318;
 Best Local Similarity 25.8%; Pred. No. 4.3;
 Matches 59; Conservative 36; Mismatches 71; Indels 63; Gaps 14;

OY 32 VGMHYKKEKN-----KDEKKRDEENKTOEHLKEITM-----IVK--IEYKGEAVK 80
 DB 53 VVHVHPVKKENLVLELDHRMKESKKREKELRKEIIEYKNPDLIREILRLLEEGIR 112
 OY 81 KEAEKLEKVPDVEYKKAIGKIYIVDGTKHISLEALSDEKKIKDIYGDALDH 140
 DB 113 REKKDIAFKYKKEALELEFR--KPYLI-----KLRERLRKIN-----LLQ 153
 OY 141 EHYVVA--KEGYEPVLIQSSDYVENT-EKALNYYEIGKILSRDILSKINOPYOKFLD 197
 DB 154 ALYLLANVAE-----MFEKEEELKEKMERAVKT-----ILFRQNGKLOSP----- 195
 OY 198 VLTNIKNASDSGDGLFTNOLKEHPTDSV-ELEQNSNEVOEYFAKA 245
 DB 196 -IGVILKN-----DEFLP--KEPPYDILSRFLQAELEPVLEKLKA 233

RESULT 11
 TTMF_CHICK STANDARD; PRT; 284 AA.
 AC P18441;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE TROPOMYOSIN ALPHA CHAIN, FIBROBLAST ISOFORM FL.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=body wall;
 RX MEDLINE=95120819; PubMed=7820856;
 RA Fanning A.S., Wolenski J.S., Moosker M.S., Izant J.G.;
 RT "Differential regulation of skeletal muscle myosin-II and brush
 RT border myosin-I enzymology and mechanochemistry by bacterially
 RT produced tropomyosin isoforms.";
 RL Cell Motil. Cytoskeleton 29:29-45(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92084115; PubMed=1748294;
 RA Lemmonier M., Balvay L., Mouly V., Labri D., Fiszman M.Y.;
 RT "The chicken gene encoding the alpha isoform of tropomyosin of fast-
 RT twitch muscle fibers: organization, expression and identification of
 RT the major proteins synthesized.";
 RL Gene 107:229-240(1991).
 RN [3]
 RP SEQUENCE OF 189-213 FROM N.A.
 RC TISSUE=Muscle;
 RX MEDLINE=89345115; PubMed=2762137;
 RA Lemmonier M., Labri D., Fiszman M.Y.;
 RT "Chick alpha tropomyosin gene contains three sets of mutually

RT exclusive alternatively spliced exons."
RL Nucleic Acids Res. 17:5400-5400(1989).
CC -1- FUNCTION: THE FUNCTION OF TROPOMYOSIN IN SMOOTH MUSCLE AND NON-
CC MUSCLE CELLS IS NOT CLEAR.
CC -1- SUBUNIT: DIMER OF AN ALPHA AND A BETA CHAIN.
CC -1- ALTERNATIVE PRODUCTS: THE MAJOR ISOFORMS OF SKELETAL MUSCLE,
CC SMOOTH-MUSCLE, BRAIN AND FIBROBLAST TROPOMYOSINS ARE OBTAINED BY
CC ALTERNATIVE MRNA SPLICING.
CC -1- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE
CC EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.
CC -1- SIMILARITY: BELONGS TO THE TROPOMYOSIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M36336; AAA65120.1; -
DR EMBL: X57991; CAA41058.1; -
DR EMBL: X57993; CAA41058.1; JOINED.
DR EMBL: X57994; CAA41058.1; JOINED.
DR EMBL: X57996; CAA41058.1; JOINED.
DR PIR: S24401; S24401.
DR HSSP: P03069; 121Y.
DR InterPro: IPR002017; Spectrin.
DR InterPro: IPR000533; Tropomyosin.
DR Pfam: PF00261; Tropomyosin.1.
DR PRINTS: PR00194; TROPOMYOSIN.
DR PROSITE: PS00326; TROPOMYOSIN.1.
KW Coiled coil; Repeat; Alternative splicing.
SQ SEQUENCE 284 AA; 32960 MW; E94405DABD65597F CRC64;

Query Match 8.2%; Score 114.5; DB 1; Length 284;
Best Local Similarity 22.9%; Pred. No. 4.3;
Matches 52; Conservative 36; Mismatches 46; Indels 93; Gaps 9;

QY 36 VKEKKENKDNRR-----KDEKRNKTOEHLKETIMKIVIEVGEAVKKAEE 85
DB 113 LEEAKKADESERGMKVTENRAOKDEKMEIOELKE-AKHI-----AEADAKR----- 161
QY 86 KLEKVPSPDLEMYKAIGKIYVDGDTIKHISLEALSDEKKIKIDYKDALHEHYV 145
DB 162 -----YEEVARKLVITIGDLR-----AEERAEI----- 185
QY 146 AKEGEPELVYQSSSEDEVNTEKALNYYEIGKIL--SRDILSKINOPYOKFLDVLTNIK 203
DB 186 -----SESQVROLEQLRINDQTLKALMAADKYSOKEDEYEEIKVL----- 228
QY 204 NASDSGDGLFTNOLKEHPT-----DFSEFLEQNSMEVOEVRPAKA 245
DB 229 -----TDKLKAEYTRAEFAERSVYKLEKSIDLDEKVAHA 263

RESULT 12
APM2_YEAST
ID APM2_YEAST STANDARD: PRT; 605 AA.
AC P38700;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE ADAPIN MEDIUM CHAIN HOMOLOG APM2.
GN APM2 OR YHLO19C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes;
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / REB526;

RX MEDLINE=95268148; PubMed=7749194;
RA STEPP J.D., Pelliscena-Palle A., Hamilton S., Kirchhausen T.,
RA Lemmon S.K.,
RT "A late coiled sorting function for Saccharomyces cerevisiae Apm1p,
RT but not for Apm2p, a second yeast clathrin AP medium chain-related
RT protein."
RL Mol. Biol. Cell 6:41-58(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favellio A., Fulton L., Gattling S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Maridis E., Meneses S., Mouser L.,
RA Nham M., Rifkin L., Riles L., St Peter H., Treviskis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.,
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII."
RL Science 265:2077-2082(1994).
CC -1- SIMILARITY: BELONGS TO THE ADAPTOR COMPLEXES MEDIUM SUBUNITS
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: U09841; AAA83415.1; -
DR EMBL: U11582; CAB34896.1; -
DR PIR: S46833; S46833.
DR SGD: S0001011; APM2.
DR InterPro: IPR001392; Adap.comp.sub.
DR Pfam: PF00928; Adap.comp.sub.1.
DR PROSITE: PS00990; CLAT_ADAPTOR_M.1; 1.
DR PROSITE: PS00991; CLAT_ADAPTOR_M.2; 1.
KW Coated pits.
SQ SEQUENCE 605 AA; 69990 MW; 7E216B11325EE3C CRC64;

Query Match 8.0%; Score 113; DB 1; Length 605;
Best Local Similarity 23.4%; Pred. No. 13;
Matches 57; Conservative 31; Mismatches 72; Indels 84; Gaps 10;

QY 37 KEKEKNDKDNRRKDEKRNKTOEHLKETIMK-----IYKIEVGEAVKKAEE 85
DB 175 KRDKRKKRKKKKKGGKSGSKSLKSIYVKNKKGINVEYKETLRNNDTGKEAAN 234
QY 86 KLEKVPSPDLEMYKAIGKIYVDGDTIKHISLEALSDEKKIKIDYKDALHEHYV 145
DB 235 ---DELPNKGNDLY-----INGDIKRTI-----IMPISMRKKGH-----Y 267
QY 146 AKEGEPELVYQSSSEDEVNTEKAL-----NYYEIGKILSDIL 185
DB 268 AKN--EFFLVYERYOYILMDKEGVIRKNLIGETVRCRYLSGMPKLKISINKILNRDPQ 325
QY 186 SKINOPYOK--FLDVLTNIK-----KNASDS-----DGQDLFTNOLKE 221
DB 326 FMSNSSFHCVSLSISINTIEKDEKNSDDAGLAQATDARIEFTIPPDGEFVLCQYELKR 385
QY 222 HPD 225
DB 386 HVXD 389

RESULT 13
IF2_AQUAE
ID IF2_AQUAE STANDARD: PRT; 805 AA.
AC O67825;


```

DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSLATION INITIATION FACTOR IF-2.
GN INB OR AQ_2032.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RM [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5.
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC -I- FUNCTION: IF-2, ONE OF THE ESSENTIAL COMPONENTS FOR THE INITIATION
CC OF PROTEIN SYNTHESIS IN VITRO. PROTECTS FORMYL METHIONYL-TRNA FROM
CC SPONTANEOUS HYDROLYSIS AND PROMOTES ITS BINDING TO THE 30S
CC RIBOSOMAL SUBUNITS. IT IS ALSO INVOLVED IN THE HYDROLYSIS OF GTP
CC DURING THE FORMATION OF THE 70S RIBOSOMAL COMPLEX (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -I- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000769; AAC0794.1; -
DR InterPro: IPR000795; GTP_EFTU.
DR InterPro: IPR000178; IF2.
DR Pfam: PF00009; GTP_EFTU; 1.
DR Pfam: PF02131; IF2; 1.
DR PRODOM: PD186100; IF2; 1.
DR PROSITE: PS01176; IF2; 1.
KW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 309 458 G-DOMAIN.
FT NP_BIND 315 322 GTP (BY SIMILARITY).
FT NP_BIND 362 366 GTP (BY SIMILARITY).
FT NP_BIND 416 419 GTP (BY SIMILARITY).
FT SEQUENCE 805 AA; 91600 MW; 8A1BF300C69CA51E CRC64;
SO
Query Match 8.0%; Score 113; DB 1; Length 805;
Best Local Similarity 25.2%; Pred. No. 18;
Matches 57; Conservative 44; Mismatches 75; Indels 50; Gaps 11;
OY 36 VKEKKNKDENKRRDEER-NKTQEEHLKEIM-----KHIVK-----IEVKG 76
DB 106 VEEIEKKKEEKEEKKEPKSVSEELIKELKEKKEKKKVEKKEKKEKRVAVYVKKE 165
OY 77 EAVKEAAEKLEKTPSDYLEMTKALIGCTIYVDGDITKHISLEALSSEKKKIKIDYGD 136
DB 166 ERKEKEKEKEKEKEPKR--IKMSKR-----EREIMR--KLEHAVKEKKKOEKREKE 213
OY 137 ALIHEHYVAKGYPVAVIQSESDYVENTEKALNVYE--IGKILSRDILSKINOPYOK 194
DB 214 KKKKEEV--KIITYIPEVIT-----VRELALLDVPANKVYALIELMKRGLVATINQVP 265
OY 195 FLVDIVNTIKMASDSGDQLFTNQLKEHPTDSVEFLQNSNEVOE 240
DB 266 EVAV-----EVAESFG-----YLAEVKKEEELKEEALLKEEBERER 302
RESULT 14
KINH_DROME

```

ID	NAME	STANDARD	PROT	975 AA
AD	KINH_DROME			
AC	P12710:			
DT	01-AUG-1990 (Rel. 15, Created)			
DT	01-AUG-1990 (Rel. 15, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	KINESIN HEAVY CHAIN.			
GN	KHC OR KIM.			
OS	Drosophila melanogaster (Fruit fly).			
CC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
CC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
CC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=89168428; PubMed=2522352;			
RA	Yang J.T., Laymon R.A., Goldstein L.S.B.;			
RT	"A three-domain structure of kinesin heavy chain revealed by DNA			
RT	sequence and microtubule binding analyses.";			
RL	Cell 56:879-889(1989).			
RN	[2]			
RP	MUTAGENESIS.			
RX	MEDLINE=93030741; PubMed=1384131;			
RA	Gho M., McDonald K., Ganetzky B., Saxton W.M.;			
RT	"Effects of kinesin mutations on neuronal functions.";			
RL	Science 258:313-316(1992).			
CC	-1- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING			
CC	PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT.			
CC	-1- SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT			
CC	CHAINS.			
CC	-1- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-			
CC	TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF			
CC	KINESIN (IT HYDROLYZES ATP AND BINDS MICROTUBULE), A CENTRAL			
CC	ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN			
CC	DIMERIZATION, AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH			
CC	INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),			
CC	VECTILES AND MEMBRANOUS ORGANELLES.			
CC	-1- MISCELLANEOUS: MUTANT FLIES DISPLAY IMPAIRED ACTION POTENTIAL			
CC	PROPAGATION AND NEUROTRANSMITTER RELEASE AT NEUROSCULAR			
CC	JUNCTIONS, BUT ARE STILL CAPABLE OF TRANSPORTING CERTAIN			
CC	MEMBRANES, INCLUDING SYNAPTIC VESICLES, TO THE NERVE TERMINAL.			
CC	-1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN			
CC	SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: M24441; AAA28652.1; -			
DR	PIR: A31497; A31497.			
DR	HSSP: P56536; 2KIN.			
DR	FLYbase: FBgn0001308; Khc.			
DR	InterPro: IPR001752; kinesin.			
DR	Pfam: PF00225; kinesin; 1.			
DR	PRINTS: PR00380; KINESINHEAVY.			
DR	SMART: SM00129; Kisc; 1.			
DR	PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.			
DR	PROSITE: PS00067; KINESIN_MOTOR_DOMAIN2; 1.			
KW	Motor protein; Microtubules; ATP-binding; Coiled coil.			
FT	DOMAIN 1 334 MECHANOCHEMICAL (MOTOR) (BY SIMILARITY).			
FT	DOMAIN 335 931 COILED COIL.			
FT	DOMAIN 932 975 GLOBULAR.			
FT	DOMAIN 180 321 MICROTUBULE-BINDING.			
FT	NP_BIND 92 99 ATP (BY SIMILARITY).			
SO	SEQUENCE 975 AA; 110429 MW; 9966C8C35BA74FD6 CRC64;			

```

QY 30 GDVGMHAKKE---KNKDNKRKDEERKNTQOEHLKIMKHIVKIE-----VKGEPAVKE 82
DQ 654 GEYRLISQHEARKMSLOESMRERANKRTLEEOJDSLREBCAKKAHEHSAVNAEKKO 713
QY 83 AAEKLEVPDVLWEMKAIKGIYIVDGTIKHIS-----LEALSEDKKIKIDYCKDA 137
DQ 714 RAEELSRMFDSQMEPLREA-----HTROVSELRDLAKOHMEKMDKHQKLL 762
QY 138 LLMHRYVAKGEPVPLVTOSSSEDEVTE-----KALN 171
DQ 763 LAHQQM---ADYEV---ROEDAEKSSSELONTILNREERQARKDLKGLDTYAKELQ 815
QY 172 VYVEIGRLSLRDLKINQPYQKFLDVNTKNASDSGDGLLTNOLKEHPTFSVEFL 231
DQ 816 TLHMLRKLFLVODLQQRIRK-----NVNNESEEDGSLAQKQ-----ISFL 857
QY 232 EONSNEVEQEV 241
DQ 858 ENNLDLTKV 867

```

```

RESULT 15
DYNAL_DROME STANDARD; PRT; 1265 AA.
ID DYNAL_DROME STANDARD; PRT; 1265 AA.
AC P13496; Q9YUAI;
DT 01-JAN-1990 (Rel. 13, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 150 KDA DYNEIN-ASSOCIATED POLYPEPTIDE (DP-150) (DAP-150) (GLUED
PROTEIN).
GN GL OR CG9206.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R, AND CANTON-S;
RX MEDLINE=87317680; PubMed=2819881;
RA Swarcop A., Swarcop M., Garen A.;
RT "Sequence analysis of the complete cDNA and encoded polypeptide for
the Glued gene of Drosophila melanogaster."
RL Proc. Natl. Acad. Sci. U.S.A. 84:6501-6505(1987).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wen K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cusack S., Dahlke C., Davaport L.B., Davies P.,
RA DePablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson R., Dou P.L., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertler S., Fleischmann W.,
RA Foster K.C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glisick A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hattis N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jalali B.E., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., Moperson D.,
RA Mervinov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,

```

```

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nussken D.R., Pacled J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000)
CC -!- FUNCTION: REQUIRED FOR THE CYTOPLASMIC DYNEIN-DRIVEN RETROGRADE
CC MOVEMENT OF VESICLES AND ORGANELLES ALONG MICROTUBULES. DYNEIN-
CC DYNACTIN INTERACTION IS A KEY COMPONENT OF THE MECHANISM OF AXONAL
CC TRANSPORT OF VESICLES AND ORGANELLES.
CC -!- SUBUNIT: LARGE MACROMOLECULAR COMPLEX OF AT LEAST 10 COMPONENTS.
CC P150(GLUED) BINDS DIRECTLY TO MICROTUBULES AND TO CYTOPLASMIC
CC DYNEIN.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.
CC -!- SIMILARITY: STRONG, TO OTHER SPECIES DYNEIN 150 KDA SUBUNIT.
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
CC FRAMESHIFTS AT POSITIONS 32, 174 TO 220, 648 TO 672 AND 1208.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, J02932; -; NOT_ANNOTATED_CDS.
DR EMBL, AE003536; AAF49788.1; -.
DR PIR, A28313; A28313.
DR FLYBASE: FBgn0001108; GI.
DR InterPro: IPR000938; CAP-Gly.
DR Pfam: PF01302; CAP-Gly; 1.
DR PROSITE: PS00845; CAP-Gly; 1.
KW Motor protein; Microtubules; Dynein; Coiled coil; Cytoskeleton.
FT DOMAIN 27 69
FT DOMAIN 105 138
FT DOMAIN 213 570
FT DOMAIN 812 836
FT DOMAIN 967 1084
FT DOMAIN 1128 1160
FT DOMAIN 1128 708
FT CONFLICT 708 708
FT CONFLICT 875 875
FT CONFLICT 888 888
FT CONFLICT 1043 1043
FT SEQUENCE 1265 AA; 141217 MW; 2038A200282B2755 CRC64;

```

```

Query Match 8.08; Score 112; DB 1; Length 1265;
Best Local Similarity 24.46; Pred. No. 35;
Matches 66; Conservative 48; Mismatches 111; Indels 46; Gaps 12;

QY 26 AGCHGVGMHAKKEKNKDNKRKDEE--RNKTOEHLK-EIMKHIVKIEKGEPAVKE 82
DQ 370 SCGGDSGSLSTYKQKQLEONIRKETLVLRDLSADKHDIOLSKLEKRESEVELE 429
QY 83 -AAEKL-----LEKVPDVLWEMKAIKGIYIVDGTIKHISLEALSEDKKIKI--KDIY 133
DQ 430 RTEKELSAKIDELAIYADVLOEQVDALGADEWEQALAEKKMEL---EDKVKLEEIA 485
QY 134 GKALLHEHRYVAKGEPVPLVTOSSSEDEVTEENTKALNYYE---IGKILSRD-ILSK 188
DQ 486 QLEALHEHRYVAKGEPVPLVTOSSSEDEVTEENTKALNYYE---IGKILSRD-ILSK 188
QY 189 NQPYQKFLDVNTKNASDSGDGLLTNOLKEHPTFSVEFL EONSNEVEQEVAFKAFAY 248

```

Mon Dec 3 10:23:30 2001

us-09-747-521-2_copy_9_282.rsp

Page 11

```

Db      544 RELVOKLNDLTLELRDRNSSNEKESLODDPSLK-----? : : : : | : : | |
          : || | : : : : : | |
Oy      249 -----YIE-----POHRDVLQLYAPAEAF 266
          : : : : : | : : | |
Db      594 TRAIIDVOLRQETELSQANHEHVMQLTAFPEEST 624
          : : : : : | : : | |

```

Search completed: December 2, 2001, 13:51:49
Job time: 344 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:48:11 ; Search time 112.1 Seconds

(without alignments)
514.086 Million cell updates/sec

Title: US-09-747-521-2_COPY_1_778

Perfect score: 3987

Sequence: 1 MNIKKEFIKIVSMCLVTAL.....TSYGRNEAFPAFRLMH 778

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3987	100.0	809	22	AA847305 Wild type B. anthr
2	3825	95.9	776	15	AA860178 Lethal factor of B
3	1322	33.2	485	22	AAU00222 Lfn-Bcl-XL apoplos
4	1309	32.8	472	15	AA860181 LF(1-254)-TR--PE
5	1307	32.8	508	15	AA860182 LF(1-254)-TR--PE
6	1306	32.8	456	15	AA860180 LF(1-254)-TR--PE
7	485	12.2	800	11	AA804236 Adenyl cyclase gen
8	253	6.3	1979	21	AA818171 Plasmodium falci
9	228.5	5.7	1786	18	AAW24790 P. falciptarum live
10	224.5	5.6	1558	21	AA818324 Plasmodium falci
11	219.5	5.5	2663	22	AA839097 Human polypeptide

12	219.5	5.5	2688	22	AA840883 Human polypeptide
13	218	5.5	980	21	AA818294 Plasmodium falci
14	215	5.4	1145	22	AA882169 S. epidermidis ope
15	213	5.3	1516	21	AA818195 Plasmodium falci
16	209.5	5.3	1780	22	AA836881 Human polypeptide
17	202.5	5.1	1788	22	AA840467 Human polypeptide
18	202	5.1	2710	17	AA895016 C. difficile toxin
19	202	5.1	2710	17	AA868387 Clostridium diffic
20	200	5.0	2485	21	AA818172 Plasmodium falci
21	199.5	5.0	1254	11	AA807503 Merozoite apical-e
22	199.5	5.0	1254	18	AAW24575 Human protein sequ
23	198.5	5.0	789	22	AA895460 Human 160KD mediat
24	198.5	5.0	1427	12	AA810534 B. burgdorferi ant
25	198	5.0	497	20	AA820046 B. burgdorferi ant
26	196	4.9	481	20	AA820047 B. burgdorferi ant
27	195.5	4.9	630	18	AAW39166 Mouse RHAMM protei
28	195.5	4.9	1392	20	AA806999 Kestlin protein seq
29	194.5	4.9	3248	17	AA899795 Kinetochore protei
30	193	4.8	1316	22	AA827248 Maize RAD50. 2ea
31	191.5	4.8	1164	17	AA885781 Group B Streptococ
32	191.5	4.8	1164	19	AAW40537 Amino acid sequenc
33	191.5	4.8	1164	21	AAW84459 H. pylori cytoplas
34	191	4.8	2440	18	AAW20828 Mutant C-beta prot
35	190.5	4.8	1164	19	AAW40541 Rattus norvegicus
36	190	4.8	1886	19	AAW54241 Human RHAMM protei
37	189	4.7	725	18	AAW39165 Plasmodium falci
38	188.5	4.7	558	21	AA818273 Mutant C-beta prot
39	188.5	4.7	1093	19	AAW40540 Human SCP-1 mutai
40	188	4.7	976	22	AA866581 B. burgdorferi ant
41	188	4.7	1087	22	AA819935 S. epidermidis ope
42	188	4.7	1119	20	AA819934 Human polypeptide
43	188	4.7	5024	22	AA882935 Mutant C-beta prot
44	187	4.7	1213	22	AAW40016 Human polypeptide
45	186.5	4.7	1099	19	AAW40538

ALIGNMENTS

RESULT 1	
AA847305	AA847305 standard; Protein; 809 AA.
ID	AA847305
AC	AA847305;
XX	
DT	29-AUG-2001 (first entry)
XX	
DE	Wild type B. anthracis lethal factor.
XX	
KW	Lethal factor: LF; immunogen; LF4: protective antigen; PA: DNA vaccine;
KW	humoral; cell-mediated; immune memory response.
XX	
OS	Bacillus anthracis.
XX	
FT	Key
FT	Peptide
FT	Location/Qualifiers
FT	1..33
FT	/Label= Signal peptide
FT	/note= "Not given in the specification"
FT	34..809
FT	/Label= LF
FT	42..285
FT	/Label= LF4
XX	
PN	WO200145639-A2.
PD	
XX	28-JUN-2001.
XX	
PF	21-DEC-2000; 2000WO-US34912.
XX	
PR	22-DEC-1999; 99US-0171459.
XX	
PA	(OHIS) UNIT OHIO STATE RES FOUND.
PA	(GALL/) GALLOWAY D R.

Query Match	Best Local Similarity	100.0%	Score 3987	DB 22	Length 809
Matches 778	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
0Y	1	NNIKKEFIKIVSMGSLVLAITSLTSGVFPIRVLQVAGHGHDVGNHVKKEKNRKNRKEE	60		
Db	1	mltkkefikivsmsectvcatltsprfiprlvgagghdvgmhnvekeknkdenrkdde	60		
0Y	61	RNKTOEHLKEIMKHIVKIEVKGEEAVKKEAEKLEKVPDSVLEMYKAIGKRIYVGD	120		
Db	61	rntkqeenhkemkhivkvevkeavvkeaeekllekrpdsvlemykaigkriyvgd	120		
0Y	121	ITKHTLSLEASHDKKKIKDIYGDALLENHVVYAKEGEPRVLIQSSDEYVNTKALNV	180		
Db	121	lthkhtlsleashedkkkkikdiygkdaillnehvyaageyprvliqssedyventekaln	180		
0Y	181	YVEICKILSRDLSKINOPYOKFLDVLNTIKNASSDSGDGLFTQOLKENHPDSFVELE	240		
Db	181	yyeigkilsrdlskinqingpyqkflavlnltiknaassdsgdglftngkhepdsfavefle	240		
0Y	241	QNSNVOEVFAKAFYVIERPQHRDVLQLYAPAFYVMKFNQEOELNLSLEELKDOQMSR	300		
Db	241	qnsnevgevfafaayierpqhrdvlqlyapaaftymdkfneqelnlsleelkdoqmistr	300		
0Y	301	YEKMEKIKOHVOMHSDSLSEEGRGGLKKLQPIEPKKODIISLSOEKELLKRIQIDSS	360		
Db	301	yekmekikohvowhmsdsslseegrggllkkqlripkpkddihsisgeekellkriqidss	360		
0Y	361	DFLSTEEKEFLKKLQIDIRDSLSEEEKELNRIQVDSNPILSEKEKEFIKLKLDIQPYD	420		
Db	361	dflsteekelfkklqidirdslseeekeellnriqvdsnpilsekekefiklklldiqpyd	420		
0Y	421	INORLODVGGLIDPSINLDVKKYKRRDIOINIDALLHOSIGSTVKNKIYLVENNMNINLT	480		
Db	421	inqrldvggllidpsinldvkkqkrtdqndalillhgsigstlvknkiyivenmnnlt	480		
0Y	481	ATLGADLVDSDTNRTINRGIFENEFKKNFYKSISSVYMIYDINERPALNBERLKWRIOISP	540		
Db	481	atlgadlvdstdnrtinrgifnefkknfykssissvymiydineralnberlkwrliqisp	540		
0Y	541	DTRAGYDENGKILLDORNTGLEIKDVQIITKOSKEVIRIDAKVYPRSKIDTKIQEQMLNIN	600		
Db	541	dtragydenktilldorntgleikdvqiitkoskeviriidakvyprrskidtkiqeqmlnin	600		

```

D6 541 dtragylengklllgrnlgleikdvqllkqsekeylridakvwpyskldtlkqaeqln 600
Q7 601 QEMKAKLGLPKRYTKITITNNVHNRVNSNIVESXVLLNEMKNKNOISDILKKYNNYLVDCNG 660
D6 601 qemakalglpkrytkititnnvhnrvnsnivesaylllnemknidqsdllkkyvnlvdcng 660
Q7 661 RFVFEDILPNIAEQYTHODEIYEQVHNSKGLVPEPSRSITLLGSPKGYELRNDSEGFIE 720
D6 661 rfvfedilpniaeqythodeiyeqvhnsgkglvypesrsitllhgspkyelrndsegfie 720
Q7 721 FGHAVDDYAGVLLDKNSDLYTNSSKFFIDIEKEBSNLTSGRTNDAEFPAEFLMH 778
D6 721 fghavddyagvllldknsdlytnsskffidiekegsnltsgyrtneaeffaeafmlh 778

RESULT 2
AAR60178
100 AAR60178 standard; Protein; 776 AA.
XX
XX AAR60178;
XX
XX 03-APR-1995 (first entry)
XX
XX Lethal factor of Bacillus anthracis.
XX
XX Anthrax; Bacillus anthracis; fusion protein; lethal factor;
XX protective antigen; cell killing; targeting; targeting; pathogen;
XX intracellular; HIV; human immunodeficiency virus; toxin.
XX
XX Bacillus anthracis.
XX
XX MO9418332-A.
XX
XX 18-AUG-1994.
XX
XX 14-FEB-1994; 94WO-US01624.
XX
XX 12-FEB-1993; 93US-0021601.
XX 25-JUN-1993; 93US-0082849.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Anora N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;
XX
XX WPI: 1994-279753/34.
XX N-PSDB; AA070179.
XX
XX Nucleic acid encoding anthrax toxin fusion protein - useful for
XX targeting toxin to specific cells, eg for killing tumour cells
XX or HIV-infected cells
XX
XX Disclosure: Page 75-77; 124pp; English.
XX
XX The sequence encoding the lethal factor of Bacillus anthracis may be
XX used in the construction of a nucleic acid which encodes a fusion
XX protein comprising the anthrax protective antigen binding domain of
XX the native anthrax lethal factor and a sequence encoding an activity
XX inducing domain of a second protein. The fusion proteins are useful
XX for the specific killing of tumour cells or the killing of cells
XX infected with intracellular pathogens, especially HIV.
XX
XX Sequence 776 AA;

Query Match 95.9%; Score 3825; DB 15; Length 776;
Best Local Similarity 100.0%; Pred. No. 8,1e-226;
Matches 745; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

34 AGGHDVGMHHVKKKKNDKNDKRDKEENKRTDEEHLKETIMKHIVKLEVGAEAVKKEAAE 93
1 agghdvgmhvkkknndkndkrdkeenkrtdeehlketimkhivklevgaeavkkeaae 93
94 KLEKVPSPDLEMYAKGKIYVDGDTIKHISLEALSDEKKKKIKDIYKDALLHEHYV 153
klekvpspdlemyakgkiyvdgdtikhislealsdekkkkikdiykdallhehyv 153

```

|||||
Db 61 kllkxpsdvlemykaigklyivdgdltkhtslslealsedkklkldygdaillhenyvy 120
QY 154 AKGYEPVLIYIOSSSEDEVTEKALNYYVEIGKILSRDLISKINQRYOKRLDVLNITKNA 213
Db 121 akegyepvliyqssedvtekalnvyeylglkilsrdlisklnpqyqkldvlnitkna 180
QY 214 SDSGDGLFTFNOLKEHPTDFSVFLEONSNEVOEFAKAFAYIIEPOHEDVQLYAPEA 273
Db 181 sdsdgdglftfnqlkhepdtfsveflqegnsnegevafakafayiepqhndvqlyapea 240
QY 274 FNYMDFNEQEIINLSLEELKDQRMLSRYEKWEKIKOHYQWMSDSLSEEGRLKKLQIPI 333
Db 241 fnymdkfnegeinlsleelkdgqrmilsryekwekikqhyqwsdlsseegrlkklqip1 300
QY 334 EPKKDIIHISLQSEKELKRIQIDSSDFLSTEFKEFLKRLQIDIDISLSEEEKELLNRI 393
Db 301 epkddiilnlsleqekellkrlqidssdfsteekeflkklqididsiseekeellnrl 360
QY 394 QVDSNPULSEKEKEFLKRLQIDQPYDINQRLQDTGGLIDSPSINTDVRKQYKRDIONID 453
Db 361 qvdsnpulsekekeflkklqidpydinqrlqidtggllidspsinldvrkqykrdionid 420
QY 454 ALHOSIGSTLTKITLYENMNINLTLATGADLVDSDFNTKINRGIFNEFKKNFYKYSIS 513
Db 421 allhgsigstlytkitylenmnlnltatlgadlvdsdfntckinrgifnefkknfkyys 480
QY 514 SNVMYIDINERPALDNERIKMRTQISPDPRAGYLENGKILILONIGLEIKDVOIINQSEK 573
Db 481 snmyivdlnierpaldnerikwrtqispdpragylengkllilgnlglekdvglkqyse 540
QY 574 EYIRIDAKVVPKSKIDTKIOEAOINIOENMKALGLPKYTKITTFPVNHNRYASINYESAV 633
Db 541 eyiridakvvpkskldtkqgeaqlndgenmkalglpkkykltitfvnhyasinyesav 600
QY 634 LILNEKNNIQSDLIKKTNYLVDSNGRVEFTDITLPNIAEQYTHODEIYEQVHSGLYV 693
Db 601 llnneknniqsdlikkvtnylvdsngrfvftdiltlpniaeqythdeleyeqvhsqlyv 660
QY 694 PPSRSITLLGSPKGVLELRNDSSEGFHIEFGHAVDVAGYGLLDKQSDLYNMSKFFIDIFKE 753
Db 661 ppsrsitllgspkgyvelrndsegfihiefghavdvagylldknsdlyvnskkffidlfke 720
QY 754 EGSNLTSGRTNEAEPFAEAFRLMH 778
Db 721 egsnltsygrtneaeffaefrlmh 745

RESULT 3
AAU00222
ID AAU00222 standard: Protein: 485 AA.
XX
AC AAU00222:
XX
XX
DT 31-MAY-2001 (first entry)
XX
DE LFn-Bcl-XL apoptosis-modifying fusion protein.
XX
KW Human; LFn-Bcl-XL; apoptosis; cancer; spinal muscular atrophy;
KW Anthrax lethal factor; neoplasm; tumour; hyper-proliferation;
KW Alzheimer's disease; neurodegenerative disorder; stroke;
KW transient ischaemic neuronal injury; spinal cord injury;
KW Huntington's disease.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Corynebacterium diptheriae.
OS Chimeric - Synthetic.
XX
XX
PH Key
FT Location/Qualifiers
FT 5..10
FT /note= "6x histidine tag"
FT 21..276
FT /note= "Anthrax lethal factor amino acids 1 to 255"

FT Region 277..485
FT /note= "Bcl-XL amino acids 1 to 209"
XX
PN MO200112661-A2.
XX
PD 22-FEB-2001.
XX
PF 15-AUG-2000; 2000MO-US22293.
XX
PR 16-AUG-1999; 99US-0149220.
XX
PA (HARD) HARVARD COLLEGE.
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Youle RJ, Liu X, Collier RJ;
XX
DR WPI; 2001-218343/22.
XX N-PSDB; AAS00250.
XX
PT Novel fusion protein for modifying apoptosis in target cell and
PT reducing apoptosis after transient ischaemic neuronal injury, has two
PT domains which targets protein to a cell and modifies apoptotic response
PT of cell
XX
PS Claim 4: Page 64-65; 65pp: English.
XX
CC The sequence represents the amino acid sequence of LFn-Bcl-XL apoptosis-
CC modifying fusion protein comprising anthrax lethal factor (LF) sequence
CC fused to Bcl-XL. The functional apoptosis-modifying fusion protein is
CC capable of binding a target cell and integrating into or crossing a
CC cellular membrane of the target cell. The apoptosis-modifying fusion
CC protein comprises at least two domains: the DFR domain, which targets
CC the fusion protein to the target cell and the Bcl-XL domain, which
CC modifies an apoptotic response of the target cell. The fusion protein is
CC useful for modifying (inhibiting or enhancing) apoptosis in a target
CC cell, such as neuron, lymphocyte, cancer, neoplasm, macrophage,
CC epithelial, stem, tumour or hyper-proliferative cell or an adipocyte. It
CC is also useful for reducing apoptosis in a subject after transient
CC ischaemic neuronal injury, especially spinal cord injury. The fusion
CC protein may be used to treat various diseases and injury conditions
CC through inhibition or enhancement of apoptotic cellular response,
CC including neurodegenerative disorders such as Alzheimer's disease,
CC Huntington's disease, spinal muscular atrophy, stroke episodes and
CC unregulated cell growth as in tumours and various cancers. The apoptosis-
CC modifying fusion protein can be delivered effectively throughout the body
CC and targeted to selective tissue and cells.
XX
SQ Sequence 485 AA:
XX
XX
Query Match 33.2%; Score 1322; DB 22; Length 485;
Best Local Similarity 58.8%; Pred. No. 4.3e-73;
Matches 293; Conservative 30; Mismatches 49; Indels 126; Gaps 11;
QY 34 AGGHGDMVHMKKEKKNDENKRRDEENKTOEHLKEIMKHIVKIEVGEAVKKEAE 93
Db 22 agghgdmvhykkekndenkrrdeenktoehlkeimkhivklevgeavkkeeae 81
QY 94 KLEKVPDVLEMYKAIGKLYIVDGDITKHISLEALSSEDKKKIKIDYIGKDALLHENYV 153
Db 82 kllkxpsdvlemykaigklyivdgdltkhtslslealsedkklkldygdaillhenyvy 141
QY 154 AKGYEPVLIYIOSSSEDEVTEKALNYYVEIGKILSRDLISKINQRYOKRLDVLNITKNA 213
Db 142 akegyepvliyqssedvtekalnvyeylglkilsrdlisklnpqyqkldvlnitkna 201
QY 214 SDSGDGLFTFNOLKEHPTDFSVFLEONSNEVOEFAKAFAYIIEPOHEDVQLYAPEA 273
Db 202 sdsdgdglftfnqlkhepdtfsveflqegnsnegevafakafayiepqhndvqlyapea 261
QY 274 FNYMDFNEQEIINLSLEELKDQRMLSRYEKWEKIKOHYQWMSDSLSEEGRLKKLQIPI 333
Db 262 fnymdkfnegeinlsleelkdgqrmilsryekwekikqhyqwsdlsseegrlkklqip1 275

```

Yy      334 EPKKDDIHSHSOEKEELKRIQIDSSPEFLS--TEKEFEFLKLQIDRIDPSIEEEKLLN 391
Db      276 -----smgsnrelv-----vdflskksqkyswsgfsdveentleapeg--- 316
Oy      392 RIQVDSNPILSEKEKEFKLKLKDIPYDINQRLQDOTGGLIDSPSIN-----LDYRK 443
Db      317 -----teseme-----tpealing--npshladsapavngatahsaldare 355
Oy      444 -----QYKRDIQNIDLHQSIST-----LNKITYLYNNMIN 477
Db      356 vlpmaavkqlreagdefeltryrafdscltsqhltpdaysfegvvneif-rdgvnmwg 414
Oy      478 NLTATL---GADLV DSTD 492
Db      415 rivaftfggalcvesvd 432

RESULT   4
ID AAR60181 standard; Protein: 472 AA.
XX
AC AAR60181;
XX
DT 04-APR-1995 (first entry)
XX
DE LF(1-254)--TR--PE(398-613) toxin fusion protein.
XX
KW Anthrax; Bacillus anthracis; fusion protein; lethal factor;
KM protective antigen; cell killing; targeting; targeting; pathogen;
KW intracellular; HIV; human immunodeficiency virus; toxin;
KW Pseudomonas; exotoxin.
XX
OS Bacillus anthracis.
OS Pseudomonas sp.
PN W09418332-A.
XX
PD 18-AUG-1994.
XX
PF 14-FEB-1994; 94WO-US01624.
XX
PR 12-FEB-1993; 93US-0021601.
PR 25-JUN-1993; 93US-0082849.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Avora N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;
DR WPI; 1994-279753/34.
DR N-PADB; AA070182.
XX
PT Nucleic acid encoding anthrax toxin fusion protein - useful for
PT targeting toxin to specific cells, eg for killing tumour cells
PT or HIV-infected cells
XX
PS Example 2; Page 90-92; 124pp; English.
XX
CC This sequence is a fusion protein comprising amino acid residues
CC 1-254 of the anthrax protective antigen binding domain of the native
CC anthrax lethal factor, a two residue linker and residues 398-613 of
CC a pseudomonas exotoxin A activity inducing domain of a second
CC protein. Such fusion proteins may be useful for the specific
CC killing of tumour cells or the killing of cells infected with
CC intracellular pathogens, especially HIV, depending on their
CC components.
XX
SQ Sequence 472 AA;

Query Match          32.8%; Score 1309; DB 15; Length 472;
Best Local Similarity 98.8%; Pred. No. 2,6e-72;
Matches 255; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

07	34	AGGGDDVMHNYKKEKKNDKKRDEENKROEHLKIMKHIVKIKVEEAAVKKRAAE	93
D5	4	aggnhdvgmnykkekknkdenkkirkdeemkqgeehlelmkhlyklevbseeavkkaae	63
07	94	KLLEKVPSPDLMTKATGKGIYVDGDTTKHISLEALSDEKKIKTDIYGDALLHEHYV	153
D5	64	kllklevpsdvlemykaigkkyivdgdttkhlslealsedkklkldiygdalldhehyv	123
07	154	AKKEGEPVLVIOSSBDYVENETKALNVYETHGKILTSRDISLKNIQPQKFLDVNTTKNA	213
D5	124	akegypvlyivqssedyentekalnvyetghkllsrdislknipqyqkflvdvnttkna	183
07	214	SDSDGQDLLFTNOLKEHPTDSVEFLLEONSNEVOEVRKAKAFYITPEQHNDVLOLVAPEA	273
D5	184	sdsdqgdlldftnqlkchpdtfsveflleonsnevgvrfakafayitpeqhdvdlolvapea	243
07	274	FNMYMKNFNEOENLSTLE 291	
D5	244	fnymdkfnegetlnlstrae 261	

RESULT

AAR60182 standard; Protein; 508 AA.

AC AAR60182

DT 04-APR-1995 (first entry)

LE(1-254)--TR--PE(362-613) toxin fusion protein.

Anthrax; Bacillus anthracis; fusion protein; lethal factor;

intracellular; HIV; human immunodeficiency virus; toxln,
 prodromal; cytotoxic

Pseudomonas sp.

W09418332-A-

18-AUG-1994.

14-FEB-1994; 94WC-0501824

9305-0021801
9305-0082849
25-FEB-1993: 9305-0082849
PR

U.S. DEPT HEALTH & HUMAN SERVICES

Arora N, Klimpel K, Leppä SH, Nichols PJ, Singh Y:

WPJ: 1994-279753/34

IN - F2DB; AAQ/0103

nuc.etc acid encoding
targeting toxin to s

OF HIV-INFECTED CELLS

Example 1; page 93-96; 124pp; English.

This sequence is a fusion protein comprising amino acid residues 1-354 of the anthrax protective antigen binding domain of the

a pseudomonas exotoxin A activity inducing domain of a second anchorage factor, a two residue linker and residues 362-013 of

killing of tumour cells or the killing of cells infected with protein. Such fusion proteins may be useful for the specific

components: intracellular pathogens, especially HIV, depending on clinical

Sequence 508 AA:

100%

Query Match 32.8%; Score 1307; DB 15; Length 508;
 Best Local Similarity 99.6%; Pred. No. 3.8e-72;
 Matches 254; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 34 AGGHGDVGMHVKEKKEKNDKDEERNKTOEHLKEIKKHIVKIEVKGEAVKKEAAE 93
 1 agghgdvgmhvkekekenkdenkrkdeernktgeehlkeikkhivklevgeavkkeaee 60
 OY 94 KLEEKPSDVLKEMKAIIGKITYVDGDTKHISLEALSDDKKIKDIYGDALHHEHYV 153
 61 kleeekpsdvlkemykaiigkityvdgdtkhislealsedkkikdiygkdalhhenny 120
 OY 154 AKEGYEPVLVIQSSSEYVENTERKALNYYEIGKILSRDIISKINOPYKFLDVLNTIKNA 213
 121 akegyepvlyiqssedyventekalnvyeyigkilsrdliskinopyqkfldvlnlkn 180
 OY 214 SDSGDGLLFTNOLKEHPTDFSVFLEQNSNEVOEFAKAFAYIIEPQHRDVLQLYAPEA 273
 181 sdsdgdgllyftnqlkehpdtfsvefleqnsnevgevfakafayyiepqhrdvlqlyapea 240
 OY 274 FNYMDKFNQOEINLS 288
 241 fnymdkfnqeelnlt 255
 DB

RESULT 6

AAR60180
 ID AAR60180 standard; Protein; 456 AA.

AC AAR60180;

DT 04-APR-1995 (first entry)

DE LF(1-254)--TR--PE(401-602) toxin fusion protein.

XX Anthrax; Bacillus anthracis; fusion protein; lethal factor;

KW protective antigen; cell killing; targeting; targeting; pathogen;

KW Intracellular; HIV; human immunodeficiency virus; toxin;

KW Pseudomonas; exotoxin.

XX Bacillus anthracis.

OS Pseudomonas sp.

XX WO9418332-A.

PD 18-AUG-1994.

PF 14-FEB-1994; 94WO-US01624.

PR 12-FEB-1993; 93US-0021601.

PR 25-JUN-1993; 93US-0082849.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Aroza N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;

DR WPI: 1994-279753/34.

DR N-PSDB: AAR60180.

XX Nucleic acid encoding anthrax toxin fusion protein - useful for

PT targeting toxin to specific cells, eg for killing tumour cells

PT or HIV-infected cells

XX Claim 7; Page 86-87; 124pp; English.

CC This sequence is a fusion protein comprising amino acid residues

CC 1-254 of the anthrax protective antigen binding domain of the

CC native anthrax lethal factor, a two residue linker and residues

CC 401-602 of a Pseudomonas exotoxin A activity inducing domain of a

CC second protein. Such toxin fusion proteins may be useful for the

CC specific killing of tumour cells or the killing of cells infected

CC with intracellular pathogens, especially HIV, depending on their

CC components.

XX SQ Sequence 456 AA;

Query Match 32.8%; Score 1306; DB 15; Length 456;
 Best Local Similarity 100.0%; Pred. No. 3.8e-72;
 Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 AGGHGDVGMHVKEKKEKNDKDEERNKTOEHLKEIKKHIVKIEVKGEAVKKEAAE 93
 1 agghgdvgmhvkekekenkdenkrkdeernktgeehlkeikkhivklevgeavkkeaee 60
 OY 94 KLEEKPSDVLKEMKAIIGKITYVDGDTKHISLEALSDDKKIKDIYGDALHHEHYV 153
 61 kleeekpsdvlkemykaiigkityvdgdtkhislealsedkkikdiygkdalhhenny 120
 OY 154 AKEGYEPVLVIQSSSEYVENTERKALNYYEIGKILSRDIISKINOPYKFLDVLNTIKNA 213
 121 akegyepvlyiqssedyventekalnvyeyigkilsrdliskinopyqkfldvlnlkn 180
 OY 214 SDSGDGLLFTNOLKEHPTDFSVFLEQNSNEVOEFAKAFAYIIEPQHRDVLQLYAPEA 273
 181 sdsdgdgllyftnqlkehpdtfsvefleqnsnevgevfakafayyiepqhrdvlqlyapea 240
 OY 274 FNYMDKFNQOEINLS 287
 241 fnymdkfnqeelnlt 254
 DB

RESULT 7

AAR04236
 ID AAR04236 standard; protein; 800 AA.

AC AAR04236;

DT 12-SEP-1989 (first entry)

DE Adenyl cyclase gene of Bacillus anthracis.

XX Adenyl cyclase; pertussis; protective vaccines; signal sequence

XX Bacillus anthracis.

OS EP366550-A.

PD 02-MAY-1990.

PF 25-OCT-1989; 89EP-0402949.

PR 25-OCT-1988; 88FR-0013952.

PA (INSP) INSTITUT PASTEUR.

PI Escuyer V, Duflot E, Mock M, Danchin A;

DR WPI: 1990-133988/18.

DR N-NSDB: Q04123.

XX Nucleotide sequence encoding adenyl cyclase of Bacillus anthracis

PT and derived proteins, useful in protective vaccines, also effective

PT against pertussis

XX Claim 8; 13; 23pp; French.

CC In vivo the adenyl cyclase protein is synthesised as a precursor with a

CC signal sequence. The mature protein is secreted into the periplasmic

CC space, the signal peptide having been cleaved off at the moment of

CC secretion.

CC Given sequence contains several regions of close homology with the Cya

CC enzyme of Bordetella pertussis. Antibodies to B.anthraxis adenyl cyclase

CC cross-react with the enzyme from B.pertussis, hence a vaccine against the

CC former species will also protect against infection by the latter.


```

Db 469 ----ihgikaelke-svklktietqelgem-----vdikqelldqj-----gek 507
OY 280 FNEOEINLSLEELKDOBMLSRY-----EKMEKIKOHYHOMSDSLSEE----- 321
Db 508 ynaqlesistelskekeynqykntyleinlnleketnkeytnlqnnynelmln 567
OY 322 ----GRGLKRLQIPIEPKDDI-----IHSLSQEKEKL---LKRQIDSSDPSTEER 368
Db 568 dlhmlngnktmktntstlcknqvhlnegidklnnekylngkslkselnvqindl--keek 625
OY 369 EFLKRLQIDIRSLSEEEKELNRIQOVSSNPLSEKEKEFLKRLK----- 414
Db 626 dflnngivdlinsnqj-----dlitrmeekeknleqenkykgemellrgnikssenlln 680
OY 415 DIOPVDINORL-----QDTGGLIDSPSILADVRKQKRIQINIDALLHOS 459
Db 681 deevcdlkrkkslkesemkmmkeehdkklaelkddcdvr--lremeknedklnmlkee- 737
OY 460 IGSTLYNKIYLENNMNNINLPTATLADLVSDTNTK-IRGIFNEFKKFKKYSISS--- 514
Db 738 -----yed-kin-----tlkeqnedkintlkeqnedkintlkekeyekintmkee 781
OY 515 -NYMTVDINERPALD---NERLKMRIOISPPTRAGYL-----ENGR 551
Db 782 yehkintlnegnehkintlnegnehkintmkeeyedkmtlnegnedkmslkeeyenki 841
OY 553 LILQINIGLEIKDV---QIIKQSEKEYIRIDAKVPRKSIDTNIQEQALINQ----- 601
Db 842 nglnsneklkdvneyleevdklvtldex---kkqgdkelnvahlakeheqillte 898
OY 602 -EMNRFALGPKYTKLTITFVHNRYASNIYESAYLILN-EMKNINIOS-DLIRKVTNLVDG 658
Db 899 meelkeqndkysdi-----yekyl-kiklsicmlintecddeneditrrleeyinnn 952
OY 659 NGRVFTDITLPLNIAEQYTHQDEIYEQVHSKGLYVPESRSILHGPSKVELRNDSEGT 718
Db 953 kql-----kveeekelkr-----hsfnllkskefr- 979
OY 719 HEFGHVVDDYAGYLDKNOSDLVTNSKK 746
Db 980 --fknsleekshelkknkexdlksdke 1005

```

```

XX 13-JUN-1995; .95FR-0007007.
PR 13-JUN-1995; .95FR-0007007.
XX (INSP ) INST PASTEUR.
PA (INSP ) INST PASTEUR.
PI Daubersies P, Druilhe P;
XX WPI: 1997-065464/06.
DR N-PSDB; AAT78868.
XX Plasmodium falciparum poly:peptide(s) and related nucleic acids -
PT derived from the liver stage antigen-3, useful for malaria vaccine
PR prodn. and diagnosis
XX
PS Claim 1; Fig 2A-I; 69pp; French.
XX
CC This sequence corresponds to a Plasmodium falciparum strain K1
CC pre-erythrocytic liver stage antigen-3 (LSA-3) protein. The encoding
CC gene sequence was isolated by screening a P. falciparum strain 79/96
CC library with serum from a missionary treated by prophylaxis (for strain
CC T6/96 see FR9101286). Of 20 clones isolated, clone 729S was used to
CC screen a library generated from Thai strain K1. One clone contained a
CC 6.85 kb insert including the genomic sequence AAT78867. The gene
CC comprises a 1.8 kb region encoding 3 major blocks of tetrapeptide
CC repeats (especially the amino acid sequence VEEV, VEEV, VEEI, VAPS, VAPT,
CC etc) and a 3' hydrophobic region corresponding to a
CC glycosyl-phosphatidyl- inositol membrane anchoring sequence. The
CC invention relates to new polypeptides of at least 10 amino acids derived
CC from the LSA-3 protein with the exception of the peptides AAM24791-4.
CC The LSA-3 peptides can be used to raise antibodies and as vaccines for
CC immunotherapy of malaria.
XX
SQ Sequence 1786 AA;

```

Query Match 5.7%; Score 228.5; DB 18; Length 1786;
Best Local Similarity 18.9%; Pred. No. 1.5e-05;
Matches 155; Conservative 134; Mismatches 271; Indels 259; Gaps 32;

```

OY 44 VKEKEKNKDEKRRKDEKRRKTOEHLKEIMKHVKIEVGEVAKKAEKLEKRP--- 100
Db 891 leekleelhenvlsaalenqsekekeydivl-----eevkeevaltlietvegae 942
OY 101 -----SDVLEMYKATIGKIV--IVDS-DITRKHSLEALSE--- 132
Db 943 eksantiteifenleenavenenvenaelnekvntvldkveetveisgeslennem 1002
OY 133 DKKKINDIYKDALLHEHYVAK-EGYEPVLVIQSE-----DVVENTEKAL 178
Db 1003 dkafseifdnkvqigenlltgmfrstelsivqsekvdenvenvsslldnlemmkegl 1062
OY 179 -----NRYVEIGKILSRD---ILSKINOPYOKFLVDVNTIK 211
Db 1063 lnklenisslegvetelvehveqnyvvdvdpamkqflgllneagglkemfndvfrk 1122
OY 212 NMSDSGOLLFTNOLKEHPTDSVEFLQNSNEVOEFAKAFAYIIEPOHRLVQLVAP 271
Db 1123 ses-----dvilveelkdep-----vakeveketvsilleemeenlvdlvee 1163
OY 272 EAFNYMDKF-----NEOEINLSLEELKDOBMLSRYEKMEKIK 308
Db 1164 ekedldkmldaaveesleisdsdsketeeslkdkkevslvveevgndmdesevkyleik 1223
OY 309 QHTQHW-----SDSLSEEGRL-----LKKLQIPIE-----PKKD 338
Db 1224 nmeeelmkdaveinditlskllietgelneveadllkdmekilekalsedskeildakd 1283
OY 339 DITHSLSQEKEKL-----LKRQIDS-----SDFLSTEE-----KEP----- 370
Db 1284 dtlexvleehndlttldvevelkdvveedklexvslkaleedllkevlekelseelle 1343
OY 371 ----LKKLQIDIRDSLS-----EEKEELNRIQOVSSNPLSEKEKEFLKRLKLDI 416

```


OY 652 TNYL 655
 1426 knfv 1429

RESULT 11

AAAM39097 standard; Protein; 2663 AA.

AAAM39097;

22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 2242.

Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokine; thrombolytic; drug screening; arthritis; inflammation; leukaemia.

Homo sapiens.

WO200153312-A1.

26-JUL-2001.

26-DEC-2000; 2000WO-US34263.

21-JAN-2000; 2000US-0488725.

25-APR-2000; 2000US-0552317.

09-JUL-2000; 2000US-0598042.

19-JUL-2000; 2000US-0620312.

03-AUG-2000; 2000US-0633450.

14-SEP-2000; 2000US-0662191.

19-OCT-2000; 2000US-0693036.

29-NOV-2000; 2000US-0727344.

WPI: 2001-442253/47.

N-PSDB: AA158253.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

Zhao QA, Zhou P, Goodrich R, Drmanac RT;

DR

Novel nucleic acids and polypeptides, useful for treating disorders

such as central nervous system injuries -

Example 4; SEQ ID NO 2242; 1007bp; English.

The invention relates to human nucleic acids (AA157798-AA161369) and

the encoded polypeptides (AA136642-AA42213) with nocotropic,

immunosuppressant and cytostatic activity. The polynucleotides are useful

in gene therapy. A composition containing a polypeptide or polynucleotide

of the invention may be used to treat diseases of the peripheral nervous

system, such as peripheral nervous injuries, peripheral neuropathy and

Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

utilisation of the activities such as: immune system suppression,

Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

and thrombolytic activity, cancer diagnosis and therapy, drug screening,

assays for receptor activity, arthritis and inflammation, leukaemias and

C.N.S disorders.

Note: The sequence data for this patent did not form part of the printed

specification.

Sequence 2663 AA;

Query Match 5.5%; Score 219.5; DB 22; Length 2663;
 Best Local Similarity 20.3%; Pred. No. 9, 1e-05;
 Matches 168; Conservative 150; Mismatches 285; Indels 225; Gaps 39;

OY 42 MHVKEKRNK-----DENKRRDEBRNKTQF-EHLKEIKMH-----YKIEV 81

DB 524 melkikexndldefealerktkkgemqllheisnlknkvkhyvngdlenelsakvel 583

OY 82 KGEAVAKKEAAEKLEKVPDVLVEMKALIGKLYIYDGDITKHLSLEASDEKKIK- 138

DB 584 lre---keqgkklqeyidsqklenlkm-----dis--ysleel edpkymqql 627

OY 139 -----DIYGDALHEHYVYAKEGEPVLYIOSSEDEVNTERKALNYVEIGILSR 190

DB 628 fdaetvaldekresafirsenlelkemke--latykgmendldqysglaakkmgv 684

OY 191 DIISKINQPKRFLDVNTTKNMSDSG---QDLFTNQLKEHPTDSVFE---LEQNSN 244

DB 685 dlekeqsaefnektltsll-----dqkvpkdlclnlelegkltldqkelnkeveena 738

OY 245 EVOEVF-----AKAFAYIEPQHHDV-----LQLYAPEAFNVMDF-----NEQETNLS 288

DB 739 lreevillseikslpseverlrkeiqkseeelnltse---kklifsevvhvestrvqgl 794

OY 289 LEBL---KDQRLMSREKWEKIKOHYOMSDSLSEGRGLKRLQPIPEPKDDIHS-- 343

DB 795 leelgtktkd-lattqsnystdtefnf-----ktclmndfeekymvleene 841

OY 344 -LSQEREKLLKRIQ-IDSS-DPLSTE-----EKEFLKQLQDIRD 380

DB 842 rmngelvnlskaeqkfidsigalktelsyktgelqektrevegerlnemeqlke-qlenrd 900

OY 381 S---LSEEEKELNRIOVDSNPLSEKEFEKRLKIDPIYDINRLQDT----- 428

DB 901 splqtverektl-----lteklqtleevklqlqekdklqqlqeslqterdqlk 949

OY 429 GGLIDSPSINLDVRYKQRYKDIONDALHOSIGSTLYLNKLYLYENMNINMLTATGLADLV 488

DB 950 sdldvtrnmndtqgeqlrnaleslqg--hgetlntlksksevsrnlh-----me 998

OY 489 DSTDNKIRNGIFNEFKNRYSSNYMTYVDINERPALDNERLKNRIQSPTRAGYLE 548

DB 999 engetek-----defgk-----mvgldkkqgdlakn----- 1025

OY 549 NGKLLIQRNTGLREKDVQIIRKQSEKRYIRIDAVVPPKIDTQIOF-AQLINQENMKAL 607

DB 1026 -----tqltadvkdelleqgrkflslqeknelqgmlesvlaegekltoldken--- 1076

OY 608 GLPKYTKLITFNVHNRYSNIVESAYLLINEMKNNTIOSDLIKVTNYLVPGNGRFVFTDI 667

DB 1077 -----iemtleng-----eeltllgdelkk--qgelvaqekhaikkegelstrcd 1120

OY 668 TLPNIAEQYTHODEIYEQVHSKGLYVPESSILHGPSKQVELRNDSEGTIHEFGHAYVD 727

DB 1121 rlaeveekkleksqgqkqqlnvqgeemseqkklneienlknlelknleltlehmte 1180

OY 728 --YAGYILDKNQSDL--VTNSKFI---DIFKEGNSLTSYGTNTA 767

DB 1181 rlelaqklneuyeovsktkekvlelqkfsteterdhlygrylea 1228

RESULT 12

AAAM40883 standard; Protein; 2688 AA.

AAAM40883;

22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 5814.

Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS;

[illegible][illegible]

XX WPI: 2000-365347/31.
DR Proteins encoded by chromosome 2 of the human malarial parasite,
XX Plasmodium falciparum, useful as anti-malarial vaccines and in the
PT diagnosis of P. falciparum infection.
PS Disclosure: Page 362-365; 577pp: English.
XX
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAH70078 to AAH70287 and AAH18144 to AAH18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
XX Sequence 980 AA:

Query Match 5.5%; Score 218; DB 21; Length 980;
Best Local Similarity 21.2%; Pred. No. 3.2e-05;
Matches 150; Conservative 126; Mismatches 262; Indels 168; Gaps 32;

0Y 51 KQENRKRDEBRNKTOEHEIKMKHIVKIEVGEEAVKREAAKLEKVPDYLEWYKAI 110
DB 345 KQENKMHVHVKIQLKREIQN--KCTSKDIECKKEQK--IKNLEDL----- 395
0Y 111 GGRYIVDGDITKHSLSALSEBDKKIKDIYQKDALHHEHYVAKGEYFVLVIOSEEDY 170
DB 396 -----ekkkienl--kdelin-----ikkmedkmmthmmdl 427
0Y 171 VENTKALVYVEIGKILSRDLISKINOPYQKFLDVLTIKNASDSGDGLFTNOLKEH 230
DB 428 lankveeln-----rlnktyekniVELn--neldvIkKKkIndeeflke- 468
0Y 231 PPDFSVFLEQNSNEQVEFAKAFAYIERQHDVQLYAPEAFNMDNFDEINLSL- 289
DB 469 -----eeekknldmwy-kIkeyeIQike-----neidslkknegnlhvl 508
0Y 290 ---EELKDQPMIL--SRYEKWEIKQHOHMSDSLSEGRGLKLRKIQPIPEPKDIIHLSL 345
DB 509 kneelnekeiIlknkydk--elnmltegyknkIgeekdmInnkIkismdqtHknq-leeemg 555
0Y 346 QEEKELLRKIQ-----IDSSDFLSTEERKFLKRLQIDIRDSLSEBEKELLNRIO--VDS 398
DB 566 eenkkelkrlknvcdmnlqsgllIkenekhmeqke-eyknlIkqdgdknllngydeR 624
0Y 399 NLFSEKEXEFL-----KKLKLDPIDYDINORQODTNGGLIDSPSINIDVRK 443
DB 625 IeIqgkemeIdyvndceekIkqakInnkIttctmnnnnmmndenIkedkkin-dlmk 663
0Y 444 QXKRDIONIDL-----HOSIGSTLYNKIYENMNIINLTATLGDADVSTONTK 496
DB 684 dnekkkeekInkIveekskIehshv--kIqemslIvegn-ekIkemgslrlIkmeel 740
0Y 497 NNGI--FNEFKKNFYKYSISNTMYIDINERPALDNERLKWRIQLSPDRAGYLENGKLI 554
DB 741 kdmekyeekkkneeerKkneerKkneerKkneekkkne--eerKkneekkklek 798

0Y 555 QNRNGLEIKD-VOLIKOSEKEYIRIDAKVYPKS-----KIDPKIOEAOINQEW- 603
DB 799 dknqfeekermeyeqkedrkrkdKkkgshsdkeekynkKekctkessnllfdeeyl 858
0Y 604 -----NRALGIPKTKTKITTFENVHNRVYS-----NYESAVLLINEW 639
DB 859 IqleelrdtgencfiyIksIskeIdv-lInkIkskdallIndafnkInlatswnlfn- 916
0Y 640 KNIQSDLIKRVYNYLVGNGRFVETDITL-PNIAEQYTHODEIYE 684
DB 917 -enkegnlttventategn-----Ittdentevemneeyk 954

RESULT 14

AAH82169
ID AAG82169 standard; Protein; 1145 AA.

XX AAG82169;

XX 03-SEP-2001 (first entry)

XX S. epidermidis open reading frame protein sequence SEQ ID NO:1432.

XX Staphylococcus epidermidis SRI strain; Infection; diagnosis;

KW vaccination; endocarditis.

XX Staphylococcus epidermidis.

XX WO200134809-A2.

XX 17-MAY-2001.

XX 09-NOV-2000; 2000WO-US30782.

XX 09-NOV-1999; 99US-0164258.

XX (GLAXO) GLAXO GROUP LTD.

XX kImmerly WT;

XX WPI: 2001-316495/33.

DR N-PSDB; AAH53019.

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,

XX useful for vaccinating against infections, e.g. endocarditis -

XX Claim 18; Page 405; 2188pp; English.

CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG81120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce host cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 1145 AA;

```

Query Match      5.4%: Score 215; DB 22; Length 1145;
Best Local Similarity 19.2%: Pred. No. 5.9e-05;
Matches 161; Conservative 154; Mismatches 271; Indels 254; Gaps 38;

QY 41 GMHVKEKKNDEKRRDEENKTOE-----EHKIKIMKHIVKEGGEAVKKEAE 93
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 361 gffmfrkkkeegqikrlnaeltqkqvngvprplredmsfltrfink--havneklkn 418
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 94 KLEKVPDVLVEMV-----KAIGKITYVDGDIR-----KHISL 127
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 419 qeydvvpelllselyqtqtslnlytlfsdevikalnkl-----enesplfeavnhqv 475
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 128 FALSEDKKIKIDY-----KDAL-----LHEHYVAKGEPVLYQSSDYEVE 175
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 476 nelsdededyeydryelntlksltslnhkyhlyhldsdlsk--ligrteftelkq 533
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 176 KALNVYEIGILSRDLSKINOPYOKFLDVL-----NTIKNASDS-----DGO----- 219
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 534 enstaaynhkhetrhneftvstngdkraldivkvpilfdrtkqdtldtllldnqtkvg 593
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 220 -----DLFTNQLEKHPDPSVEPLEONSNEVOEYAKAFAYIEPQHRDV 265
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 594 vfgtfsagksallnallgenylvespmtaateelsygsesqtltskegllle-eynhv 652
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 266 LQLYAPAFNMDKFNQOEIN--LSLEELKQD--RMLSRKEKEKIKOHYQHMSSDSISEGR 323
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 653 lefy-eaistnldfdifesddlkikleknglafisalek-----hyemysml----- 700
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 324 GLKKLQIPPKKDDIRHSISOEKEELKRIQIDSSDFLSTEKEPEFLKQID----- 377
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 701 -----ehslhtvsleeik-----kwsadeyaltfktvhlkkipidklski 742
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 378 IRDSL-----SEEEKELNRIOVSSNPL-----SEKEPEFLKRLKIDIOPYDINQ 423
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 743 lldslglsnmgqrhtneeqqltsadllytyfnhsftndkafiehm-----dmng 796
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 424 RLQPTGGLDPSINLDVRYKRYKDIONIDALHNSIGS--TLNKKIYLY-----ENM 474
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 797 lneqatfmlinaavdl--aedkqdlqavedyvadalqgvnlhseilyvsrsqslngnl 853
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 475 NINMLTAT-----NTRKINGRINEFKKFKFSSISSNYMIVDINE 523
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 854 glnlresiaqyfvakevseattlegqmtylqgmntsfkmmidfdndnaklsarqklnhy 913
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 483 -----LGADLVDSFD-----NTRKINGRINEFKKFKFSSISSNYMIVDINE 523
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 914 knqtrlnqelldtaqtrfneveeqvylhnerlkqlldexkvsfnsmtqnn--dfne 970
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 524 RPAIDNERLKRIOQLSPTRAGVLENGKLIQR-----NIGLETQVQIIIOSEKEYIRI 578
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 971 -----ekkiskilyldqthqqlfleg-silferlkkyfnsqdeqlipvmkklnglnhvl 1024
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 579 DAKV-VPKSKIDTRIOEQALNINQEMNKALGLPKY--TKLITFNVHRNRYSNYESAYL 634
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1025 nakfnvpsvstdpqlldqeln-----smldgslpqktrkikvnpksqgdqehlangtle 1079
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 635 ILNEMKNNIQ--SDLIKRVTVNLVDNGRFEVFDITLPNIAE--QYTHODEIYEQVHSK 689
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1080 llqgdlnsltrqldnylnemtcg-laebqfgmletsig-qgidelstlftddtlldqlelk 1137
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

XX XX Plasmodium falciparum.
OS OS
XX XX
PN PN WO200025728-A2.
PD PD
XX XX 11-MAY-2000.
XX XX
PF PF 05-NOV-1999; 99WO-US26796.
PR PR 05-NOV-1998; 98US-0107131.
XX XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
XX (VENT/) VENTER J C.
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WP1; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 120-124; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAH70078 to AAH70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
SQ Sequence 1516 AA:

```

```

Query Match      5.3%: Score 213; DB 21; Length 1516;
Best Local Similarity 19.8%: Pred. No. 0.00011;
Matches 162; Conservative 136; Mismatches 271; Indels 248; Gaps 37;

QY 45 KEKEKN-----KDEKNKDEERNKTOEHLKEIM-----KHIVKEGGEAVKKE 90
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 677 klskntnlnlkkynvqkkskkkkkflndlnlytfnsteskylgqlykksge--ke 734
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 91 AAEKLEKV-----PSDVLMEYKAGIKYIVGDIRKHISLEALSDK 134
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 735 dlnqldfvtqcyrmndltdnksdflknklidmknkyl--ynlejeqeelnekk 791
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 135 K-----KINDYGDALLHEHYVAKGEPVLYIQSSDYEVEVENTEKALNV 180
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 792 nynknmdsnkftflknefekollldsqdflgslldlkeynytdndnnmneksl 851
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 181 YVEIGILSRDLSKINOPYOKFLDVLNTRKASDS--DCODLFTNQLEKHP--TDFS 235
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 852 yedgenfltr-----neplcneyeeknlllylsdegkyneedlfxkklkkekndnts 905
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 236 VEPLEONSNEVOEYAKAFAYIEPQHRDVQLVLAPEAFNMDKFNQOEINMSS--LEE 291
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```



```

Db 906 sddfence--vqe-----kiyvnekieynknndksssssiilee 945
QY 292 LKDOR-----MLSNYEKWEKIKOHYQHW-----DSLSEEGRLK 328
Db 946 lkykkekdelvspnlcvlldefehsndlenyisvssddmktnvskniltgvkenkydk 1005
QY 329 LQIPPEPKKDDIHSLSOEKELLRIOIDSSDFLSTEEKEFLKRLQIDIRDSLSEERE 388
Db 1006 tnveydkkgddgvleisfedshkleeskfdannlydnndeleknls--kdyisgvdkn 1062
QY 389 LLMRIQVDSNPLSEKEKEFLKRLKLDIQPYDINQRLQDTGGLIDSPSI--NLDVRKQY 445
Db 1063 hvnnl-ynlergederenef-----venkligsteeshkenefictenkslirkqy 1109
QY 446 --KRDIONIDALLHQSISTGLNKLTYLENMNNINLTATLGAD-LVDSTDNTKINRGIFN 502
Db 1110 mskedlsnvril-----ksddlnlksqnyfeliildk-----k 1142
QY 503 EFKKNFKYSISSNYMIVDINERPALDNRLKWRIOQSPDTRAGYLENGKLIQRNIGLEI 562
Db 1143 qymdnfmgmlegn-----ndklyk-edkldegayfeylednkil----- 1179
QY 563 KDVOIIOSEKEEYIRIDAKVVPKSKIDYKIQEAQJLINOEWNKAL-----GLPKYTKL 615
Db 1180 --dsyiketnke---neelikeyk--klkknleindemnddikllnffgipyiqsp 1230
QY 616 ITFNVHNRYASN-----IVESAYLLINEMWNNIOSDLIKKVTNYLYVDGNGRFEVFTDIT 668
Db 1231 ceaeaqcsylnnknycdalisdsvlvfsgkvtvknfknkkt----- 1274
QY 669 LPNIAEQYTHODEIYEQ--VHSK-GLVVPESRSI-LHGSPS-----KGVELRNDSGFIH 719
Db 1275 -----veyekakeekiglygeelnnislcgcdytlisvnhgigivnale-lik 1322
QY 720 EFGHAVDYAGYLLDKNOSDLVTNSKRFID--IFKEE 754
Db 1323 afpnfed-----lklldivsnpfkrldkmynee 1352

```

Search completed: December 2, 2001, 13:48:20
 Job time: 140 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OH protein - protein search, using sw model

Run on: December 2, 2001, 13:46:05 ; Search time 132.69 Seconds
(without alignments)
891.810 Million cell updates/sec

Title: US-09-747-521-2

Perfect score: 4145

Sequence: 1 MNIKKEFIKIVSMSCLYTAI.....KNAPKPEQFINDKIKIINS 809

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_17.*
2: sp_archaea.*
3: sp_bacteria.*
4: sp_fungi.*
5: sp_human.*
6: sp_invertebrate.*
7: sp_mammal.*
8: sp_mmc.*
9: sp_organelle.*
10: sp_plant.*
11: sp_protist.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	278	6.7	1946	5	097291 plasmodium
2	273.5	6.6	2752	5	09BJY0 plasmodium
3	264.5	6.4	1979	5	096133 plasmodium
4	263.5	6.4	2771	5	026216 plasmodium
5	255	6.2	2269	5	026223 plasmodium
6	254.5	6.1	2747	5	09BJX9 plasmodium
7	248.5	6.0	3130	5	09BK46 plasmodium
8	248.5	6.0	3254	5	09BK45 plasmodium
9	246.5	5.9	1939	5	025662 plasmodium
10	245.5	5.9	1127	12	09YVR6 09yvr6 melanoplus
11	245.5	5.9	1387	5	096276 plasmodium
12	239.5	5.8	3119	5	025857 plasmodium
13	238.5	5.7	1238	12	09EMP3 025857 plasmodium
14	235	5.8	1048	5	026023 plasmodium
15	229	5.5	3724	5	077320 plasmodium
16	228.5	5.5	1786	5	09U0P0 plasmodium
17	225.5	5.4	1931	5	09NCF9 09nct9 drosophila
18	224.5	5.4	806	2	09LBP7 09lbp7 mycoplasma
19	224.5	5.4	1558	5	096275 plasmodium

20	223.5	5.4	1938	13	09IBD7 09ibd7 seriola dum
21	222.5	5.4	2166	2	051465 051465 borrelia bu
22	222	5.4	2867	5	09N2M3 09n2m3 plasmodium
23	218.5	5.3	1931	5	09YKH9 09ykh9 drosophila
24	218	5.3	980	5	096246 096246 plasmodium
25	217.5	5.2	999	5	0905A3 0905a3 plasmodium
26	217	5.2	1088	5	000905 000905 oxytricha f
27	216	5.2	1005	1	058718 058718 methanococc
28	214.5	5.2	1313	10	09XIP6 09xip6 arabidopsis
29	214	5.2	3259	4	014789 014789 homo sapien
30	213	5.1	1516	5	096154 096154 plasmodium
31	212.5	5.1	1065	2	09AHK8 09ahk8 borrelia bu
32	211.5	5.1	1065	2	09AHK7 09ahk7 borrelia bu
33	211.5	5.1	1206	4	09NTC1 09ntc1 homo sapien
34	209.5	5.1	976	2	054222 054222 staphylococ
35	209.5	5.1	1128	10	09SAE6 09sae6 arabidopsis
36	209.5	5.1	1780	4	09UFR5 09ufr5 homo sapien
37	208.5	5.0	841	10	023037 023037 arabidopsis
38	208	5.0	977	5	09UOH9 09uoh9 plasmodium
39	208	5.0	1011	2	09AHL0 09ahl0 borrelia bu
40	207.5	5.0	683	2	050281 050281 mycoplasma
41	207	5.0	849	2	09AHK4 09ahk4 borrelia bu
42	207	5.0	3394	5	077384 077384 plasmodium
43	206.5	5.0	1650	5	077328 077328 plasmodium
44	206.5	5.0	1734	5	09U0L7 09u0l7 plasmodium
45	206.5	5.0	1837	3	074424 074424 schizosacch

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	1946 AA.
ID	097291			
AC	097291			
DT	01-MAY-1999 (TREMBLrel. 10, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)			
DE	HYPOTHETICAL 231.8 KDA PROTEIN			
OS	Plasmodium falciparum (isolate 3D7).			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI:TaxID=36329;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=3D7.			
RX	MEDLINE=99376085; PubMed=10448855;			
RA	Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,			
RA	Churcher C.M., Craig A., Davies R.M., Devlin K., Fellwell T.,			
RA	Gentles S., Gilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,			
RA	Horrocks K., Jagels K., Jassal B., Kyes S., McLean J., Mole S.,			
RA	Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,			
RA	Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,			
RA	Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;			
RT	"The complete nucleotide sequence of chromosome 3 of Plasmodium			
RT	falciparum."			
RL	Nature 400:532-538(1999).			
DR	EMBL: AL034559; CAB39037.2; -			
DR	InterPro: IPR002048; EF-hand.			
DR	PROSITE: PS00018; EF_HAND; UNKNOWN_2.			
KW	Hypothetical protein.			
SO	SEQUENCE 1946 AA; 231792 MW; 59AC248AB980E34 CRC64;			
Query Match	6.7%; Score 278; DB 5; Length 1946;			
Best Local Similarity	21.6%; Pred. No. 0.00075;			
Matches 189; Conservative 159; Mismatches 315; Indels 214; Gaps 45;				
OY	47 KEKNKDENRKRDEKRNKQ-----EENHKEIKMHIIVKIEVGEAEVKK 89			
DB	1051 RDNHNSNNKNNKNNNNYYYYQNNNNLSHTVLEKKKALNKQYKYLENKLIVOKK 110			
OY	90 EAA-----EKLEKVPDVLNWKYKAIGKTY---IVDGDITKHSLEALSDEKKIKID 140			

D	b	1111	EMSCNKKKKKKKSLINEYEKIKLIDNIIDENPKLKENCIRKKKKLAKMDITRYINEQ	1170
O	y	141	YGNAL-----LHEHYAAEGEPVLVIOSSDYEVTETKALNYEGLISDLISK	195
D	b	1171	FSLSKIDPENKKNYEYLFELKKRUSEIYML---KELKEKEKILYNDILKYKKADVDI	1227
O	y	196	INQYQKFLVDLNTIKN---ASDSGDGLFTNQLKEHPDPFSEVPELQNSNEQVEFA	251
D	b	1228	LKEMIEKIDIDIKKLLKQGEIISQKRQIETLENNLK-----IGKEKINKDNEIQLKQY	1282
O	y	252	KAFAYIEPQ-----HNDVQLYAPAFNYMDKFNQEIINL--SLEF-LKDGRLS----	299
D	b	1283	K-INIHKEKEIKKTDIEKEKNLTKLSKFDIINKENNMLNKLETTLLKNEKEISTENV	1341
O	y	300	-----RYEKMENTIKQHYQWMSDLSSEGRGLKILQIDPIEKK	337
D	b	1342	KLIETNKTLIYENKLSNDLKNLTKLEKEIKQENY-----NKIKKEIKITIEIQLQYKEI	1397
O	y	338	DDI-----IHSLSQEKELLRKIDIDSSDPELSTEKEFLKQIDIRDS--LSEEE	386
D	b	1398	KHMKEITQIKQIHNLDKH---IAQIQEKNMF-----DESLYKEKNEMKSNILEEY	1449
O	y	387	KEL-----LNRIOY-----DSSNPSEKEKEFLKTKLDIOPYDINOQLDTGGGLD	433
D	b	1450	KELSTYEIDKNINKIKIEDLEKEKNELTKNEE-INNLKEEYKM--VOQHLEDNYLVE	1506
O	y	434	SPSINLD-VRKQYKRDIONIDALLHOSIGSTLYNKIYLYENN-----INNLTATIGADL	487
D	b	1507	KQKLADITREKKNIIINEDCKIKNNK--KLKLNK--LKENQNNYEHTLNNI-----	1554
O	y	488	VDSDNTNKRINRGIFNEKKKFKYSIS-----NY-MIVDINERPALDNERLKWRIOLS	539
D	b	1555	-----KKNQOIIIEERKKNTQYKVSLEHAFKQSYNOQLDQNNEMLOOIKOLK--NVMD	1607
O	y	540	PDTRAGYLENGCKLLQGNICLEIKDVOIJKOSEKEYIRIDAVVPKSKIDTRIOEALNI	599
D	b	1608	IKTMSKNLKNVNNEL-----IKETKNYSQCKEKETK-GLKNIGQATIKLKNENQOLKI	1659
O	y	600	NO-EMNKALGLPKYTKLITENVNHRVASNIVESAYLILNEMKN-NIGSDLIKKYTVNLVD	657
D	b	1660	NAFEYIKKDVODNY---VTLNHNHNN-----ILNEOKRLVQIDILKS---QVD	1701
O	y	658	GNGRFVFPDITLPIALAOYHODEIYEOYHSGGLVYRESRILLGPGSGVGLRN-----	712
D	b	1702	QKQNI-----NNMKQ-----IEDVNHKIASINKEKEEL-----NTTKIKKTKTED	1744
O	y	713	--DSEGFIEFGHAVDVAGYLLDKNQSDLVYNSKRFIDIEKEGSM-L-TSYGRTENAE	768
D	b	1745	VNLSEKILSELNSKDEDEVKKRTIEIKQKE--REYKKLLDVKIEKKNLVRYEK--ELD	1800
O	y	769	FFAEAFRLMHS--TDHAERLKVQKNAPKTFQFQINDOI	803
D	b	1801	SYMTRYEFPAHAKKYQYEEETIKDLKNKLK---LKDEV	1833
RESULT 2				
O9BJYO PRELIMINARY; PRT: 2752 AA.				
AC	O9BJYO;			
DT	01-JUN-2001 (Tremblrel, 17, Created)			
DT	01-JUN-2001 (Tremblrel, 17, Last sequence update)			
DT	01-JUN-2001 (Tremblrel, 17, Last annotation update)			
DE	235 KDA RHOPTRY PROTEIN (FRAGMENT).			
GN	ITIA.1.			
OS	Plasmodium yoelii yoelii.			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=73239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	SPRAIN-YM.			
RT	Khan S.M., Jara W., Peter P.R.:			
"Distribution and Characterization of the 235 kDa Rhoptry Multigene Family within the Genomes of Virulent and Avirulent Lines of				

RT	Plasmidium	yeo11.1;
RL	Mol. Biochem. Parasitol.	0:0-0(2001).
DR	EMBL; AF323442; AAK15625.1;	-.
FT	NON_TER	1
FT	NON_TER	1
FT	NON_TER	2752
SO	SEQUENCE	2752 AA; 322606 MW; EB34E3753168AF76 CRC64;

Query Match	6.6%;	Score 273.5;	DB 5;	Length 2752;
Best Local Similarity	19.6%;	Pred. No. 0.0018;		
Matches 191;	Conservative 158;	Mismatches 283;	Indels 343;	Gaps 44

Qy	58	DEENKROEHLKIMHIAVIEVKEGEAV-----KKAARKELEKVSODLEMYKAI	110
Dp	1736	NKEMKTYNEMMEYNILVDFLEVSSEITYEIKKRIARSTOKELKSIENYK-----	179
Qy	111	GGKIYIVDGDITFKHISLEALSDEKKIKODIYGDALDHEHYVAKEGEYEV-----	161
Dp	1791	-AKSYL--DDETNEFDRIYTHFKKKINDVNDK---FTNEYSKVKNCFDINSINNVAK	184
Qy	162	-----LVIOSSEDEVENTER-----ALNYYIEIGKILS-----	189
Dp	1845	STDENLLLSILNQIKEMYANIVSKYYSYKVEANIPINISKLANSLNIOQSSGIDLH	190
Qy	190	-----ROLLS-----KINOPYKFDLVLTIKN-----ASDSOG	219
Dp	1905	KNINIAIVSYLNSOKEMLSIPSPOKTSEYTKISDSYNTLLDIFKKSXELOKKEOAL	196
Qy	220	DLLF-----TDLOKHPDFF-----SVEFLQNSNEVOEYFAKAPAYIE	259
Dp	1965	NLIEPNRLHDKVOATNELKDOTLSDLKKNKEQOILINKYXLLHKSNELNKLSGSS--	201
Qy	260	POHRDVLQVAPAFNMYDKNEEOINLSLEELK-----DORMLSRYEKWEKIKO	309
Dp	2019	-QNYDTIL-----ESSKY-DKIKESNNYBDEKKNKLGDIDPVTAMEKKFNNDIDIELEW	207
Qy	310	HYQMSD-----SLSEEGRGLL--KKIQ-----IPREKKDD-----IHSLSOEK	349
Dp	2073	NYKH-SDKDYNFSEBENNNTLOSRRKKELESMAFNAIKKXKXKIEKNOLINKLLETTR	213
Qy	350	ELL-----KRIOI-DSSDPLSTP-----EKFEFLKLOIDRDSLSSEKTEL-----L	390
Dp	2132	ECMEFTYTTLVEALRIRKITDYSKRTSATFRSKEFLYID-DTSNLSINDIDINTLOTKYDL	219
Qy	391	NRIO-----VDSNPLSEKEKEFLKRLKLDIOPYDINORLODPTGCLDPSINID	440
Dp	2191	NQIKKHTWSMEADATNDNNNLIEREKATYTIINNLTELF-----TIDSNNIDAD	223
Qy	441	VRKOYKRDIONIDLHOSIGS--TLNKKIYLENNMINNLTLATGLADVLDSIDN-----	493
Dp	2240	VLHNNKIOMIYFENSELHKSIDISIQOLKKNHVEFKLNGIHNKKY-FDISKEPDNIILOQ	229
Qy	494	-TKINRGJFNEFK-NRKYSSISSYMYLVDINERP-----ALDNERLKWRJOLSPDR	543
Dp	2299	ESELKENL-NOLUKTIGQISDKKKFELNELSELPFNEFTLKEIYHEIYVESOIDEIKN	235
Qy	544	AGYLENGKLILORNIIGLEI-----KDOVLIKROSEK	573
Dp	2358	ITNEBENDNITLYMLOITIRKMKKVSILNFPYTYENDSNVILKOHIONNENNDVSKIDNLIK	241
Qy	574	EYIRIDAKVPR-----SKIDFKIQEAOILINDEW--NRKALGPK-611	
Dp	2418	KTIESFOKITLKEIKAOFPDNNNNINNVISTISODVIVKHKISDLTYENELIETOS	247
Qy	612	--YRKLITFNVHNKYSNIVESATLLNENKKNIOS-----DLIKATYNTLVGNGR	661
Dp	2478	LEYIKRSYTDIRSEIOITRYVPHIDHYEOOTKKIÖDNPKNDEIDDLÖEIVANNKE----	253
Qy	662	FVFPDITLPN-----IAEOYHOEIEYOVSUKLYVPESRSL	700
Dp	2534	---SEKLPITLINNKDNVPTPLISIKDKVYNLIKSEYNNNDNVSYNNAKK-----	257
Qy	701	LHGSGKVELNDSGEFTHIEFGHAVDDYAGYLLDKNQ--SDLVTSKKEIDIFK-----	752

```

Db 2580 -----LEEDANSIIRD-----LDTSQNNLNDLQKNLKIIDLKKNKE 2618
Oy 753 -EESNLTSGRTNE 766
Db 2619 IENRYNLQTNREOE 2633

RESULT 3
ID 096133 PRELIMINARY; PRT; 1979 AA.
AC 096133;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE HYPOTHEMETICAL 237.7 KDA PROTEIN.
CM PEB0145C.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99021743; PubMed-9804551;
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shalton S., Mason T., Yu K., Fujil C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
DR EMBL: AE001375; AAC71819.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1979 AA; 237745 MW; 5C6CD6307AEFD37 CRC64;

Query Match 6.4%; Score 264.5; DB 5; Length 1979;
Best Local Similarity 19.0%; Pred. No. 0.0028;
Matches 169; Conservative 178; Mismatches 286; Indels 255; Gaps 40;

Oy 42 MHVKEKNKDNKRKRDEERNKTOEHLKEIKHIVKIEVKEEAVKKBAEKLLEKVS 101
Db 329 LNKOKEREKEREKEREKEREKEREKEDTLIK-----ELKDEKI-----SLEKVS 376
Oy 102 DYLEHYKAIIGCIYIVDDDIK--HISLEALSEDKKIKIDYKQALHHEHYAKEGE 159
Db 377 -----IKVREMDIEREHNFLH-MEDQKLDKKNKSVKNN--NQLVYKCE--- 418
Oy 160 PVLVIOSEEDVENTEKALNLYVEIGKILSRDILSKINQPYOKFLDVLNTIKNASDQ 219
Db 419 ----IKNKTELEKKEKELK---DIENVSKEINKLNKKEKQIILAFKNHKEE--- 468
Oy 220 DLLFTNOLKEHPTDFSVLEQNSNEVOEFKAPAYITEPOHRDVLQLYADEAFNYMDK 279
Db 469 ----IHGLKEELKE-SVKITKIELOELQEM-----VDIKQELDQ-----QEK 507
Oy 280 FNEOEINLSLELKNQRMLSRY-----EKMEIKOHYQMSLSSEF----- 321
Db 508 YNAQTESISTELSKKEKEYNKNYIEBINLNKLEETNKNEYNNLNANNYNEINMLNN 567
Oy 322 ----GRGLKQLQIIEPKDDI-----IHLSQSEKEL--LKRIQSDSDFLSTEK 368
Db 568 DIHMLNGIKTMNTOISTLKNVHLNEQIDKLNNEKGLNKSISELNVQIMDL--KEEK 625
Oy 369 EFLKQLQIDIRDSLEEKELLNRIQVSSNPLSEKEKEFLKLKL----- 414
Db 626 DFLNQIYDLISNQI-----DLTRKMEERENKMLEQENKYQEMELLRGNKISENILLN 680
Oy 415 DIQPDINQRL-----QDTGLIDSPINLADVRYKQYKRIQIWDALLHOS 459
Db 661 DEEVCDLKRKLSKESEKMKKEEHDKLAEKDCDVR--IRENNEKEDDINMLKEE- 737
Oy 460 IGSTLYNKIYLYENNNINNLATGLADLVDSQNTK-IRGIFNEFKKFKXSIS-- 514

```

```

Db 738 -----YED-KIN-----TLKEQNEDKINTLKEQNEDKINTLKEEYEHKINIMKEE 781
Oy 515 -NMYIVDINERPALD---NERLKWRIQLSPDTRAGY-----ENCK 551
Db 782 YEHKINTLNEQNEHKINTLNEQNEHKINTMKKEEYEDKMMTFLNEQNEDKMNSLKEEYENKI 841
Oy 552 LILQRIQGLEINDY--QIIKQSEKEIIRIDAKVVPKSIDTKIQEQLNINQ----- 601
Db 842 NOINSNNEIKIDVNEVEYIEVDKLVTLDEK---KKQFDKEINVAHKAHEQOILLTE 898
Oy 602 -EMNKALGLPKYTKITLTPVHNRYASNIYESAVLIIN-EMKNKIOS-DLIKKTNYLVQ 658
Db 899 MEELKQQRNKKYSDL-----YEKTI-KLKISCIIMINIECCDIEDIEDIIRIEEYINN 952
Oy 659 NGRFVFTDITLPNIAEQYTHODEIYEQVHSGKGLVPESSILLGSPKGEVLNDSGFT 718
Db 953 KGL-----KKEVEKEHKR-----HSSFNILKSKEKF- 979
Oy 719 HERGHAVDDYAGYLLDKNSDLVTNSKKF-----IDIFKEGSMN 758
Db 980 --FKNSIEDKSHELKKRHKERDLISKDETEENKKIKTELINDIKILODELVYKKO-SNA 1036
Oy 759 TSYGRTEAEFEFAEAFRLMHSSTDHAERLKYOKNAPKTFQFINQIKFI 806
Db 1037 QQVDHKKKSWILLKDKSKKIKDKENQINVEKNKEEDLKKKDEIRIL 1084

```

```

RESULT 4
ID 026216 PRELIMINARY; PRT; 2771 AA.
AC 026216;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE RHOPTRY PROTEIN.
OS Plasmodium yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5861;
RN [1]
RP SEQUENCE OF 379-2771 FROM N.A.
RC STRAIN-YM;
RX MEDLINE-97077455; PubMed-8920022;
RA Sinha K.A., Keen J.K., Ogun S.A., Holder A.A.;
RT "Comparison of two members of a multigene family coding for high-
RT molecular mass rhoptry proteins of Plasmodium yoelii.";
RL Mol. Biochem. Parasitol. 76:329-332(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-YM;
RA Green J.L., Holder A.A.;
RT "Structure of the E8 gene encoding a high molecular mass rhoptry
RT protein of Plasmodium yoelii.";
RL Mol. Biochem. Parasitol. 0:0-0(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-YM;
RA Holder A.A.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: U36927; AAB41263.3; -.
SQ SEQUENCE 2771 AA; 325640 MW; C0CCB9A6E7ACF36 CRC64;

```

```

Query Match 6.4%; Score 263.5; DB 5; Length 2771;
Best Local Similarity 20.3%; Pred. No. 0.0047;
Matches 193; Conservative 168; Mismatches 358; Indels 233; Gaps 42;

Oy 2 NIKKEIKYISCVLTATLTSPVPIPLVQAGAGGDVGMHKEKKEKKDEKRRDEER 61
Db 1759 NIYEPIKSYDILTHYLETVSKEPITYEQIKNKRITAEQELLINIKNVKAKSYLDIDIA 1818
Oy 62 NKTOE--EHKKEIMKI-----VKIEVKEE-----AKKKAELKLEKVPDVL 104

```

```
Db 1819 NEPDRIYTHKKNLNDVNDKFTNEYSKVN-KGFNINISNINNVKKSDBENLLINLQTK 1877
OY 105 EMWKAIGCKIYI-----VDG-DITKHIS---LEALSEDK 134
Db 1878 EYMANIVSKKYSYKYEAEINIFINIPKLANSLNIQIKSSSGIDLFKNINIMILPYLDSOK 1937
OY 135 KKIDYIGKQALLHENVYAKEGEYPLV-VYQSSDEVYENKALNYYEIGKILSRDI 193
Db 1938 KDTLFTIPSEPTSETYKISDSYNTLLDILKRSOELQKKEQOALNLIFF-----NRLH 1992
OY 194 SKINPOKQFLVNTIKNASDSODGLLFTNOLKEHPDTSVEFLQNSNEVOEYFAKA 253
Db 1993 DKV-QATNELKQDLSDLKKNKEQ-----ILNKTK-----LLHKSSELKLSGNS 2036
OY 254 FAY--YIEPOHADVQLYAPAFNYMDKFNQOEINLSLELKDQRLMSRYEKMEKIROHY 311
Db 2037 QNYDTILESSKYDKIK--EKSNNYEKEKEKLGINFQYKAMEQO-FNNIDIDIEKLENNY 2092
OY 312 QHMSD---SISEGRGLL---KKIQ-----IPIEPK---KDDIHSLSQEKKE-- 350
Db 2093 KHSEKDNVNFSEENNINLOSRRKLLELTNAFNAEIKRIEDIIIEKNGLINLLETTRDCM 2152
OY 351 -----LLKRIQIDSSDF-----LSTEEKFELKQIDIRDSLESEKEL-----LNRI 393
Db 2153 LFYTKTLVEPLKIKITTYTKFTSATKFSKEFLKYIDA-TSNSLNDIINTLQTYKDLNQI 2211
OY 394 Q-----VDSNPLSEKEKEFLKKLKDIOPIYDINORLODTGGLDPSINLQVRK 443
Db 2212 NKHVASVADATNDNNMLIEKEKEATKTINLTLEF-----TIDSNKIDADGLH 2260
OY 444 QYKRDIONDALHOSIGS--TLYNKIYVENMNI-----NNLTATLQADL 487
Db 2261 NKKIQTILFNSSELKSIDISIKQILKMKHAKKLLNIGHINKKYEDISKEFDNITLOLQESL 2320
OY 488 VDSNTDKINKNGJFNERKNEKFKYSSISNYMIVDINERPALDNERKRIQLSPTTRAGYL 547
Db 2321 TANLNDKEIGQKISDKKQKQHLALSET-PIPNENTLKEIYHDIYVKKNOJDELENTNE 2379
OY 548 ENKLLIQRNIGLET-----KVOIITKQSEKYEYRIDAKVYPKSIDTKI 592
Db 2380 ENENITLYMIDITTKLMKVOSILNFVTYENDSNIIKOHIODNNENYVSKI-KESLETTYI 2438
OY 593 QEOALNIOENMKALGLPKYTKLITFPVNHKVASINIVESAVALLNEMKNINOSLKKVT 652
Db 2439 QSFQKILN-----KLNKIKAOFTYDNNNNINNVISTISQDVNDVKKHISKDL--TIE 2486
OY 653 NYLVDSGRFVETDITLPIAEOY--THODEIYEQVHSGKLYPESISILHGPSKGYEL 710
Db 2487 NELIQ-----IQKSLIEDIKKSTYDIRSEQITKYVNPIDHYVQEQTKKIQNNPNK---- 2535
OY 711 RNDSEGIHERGAHVDTYAGILDKN-QSDL-----VTNSK-----KFIIDIRKEE 754
Db 2536 -----DEIDDLQIEIVYVKNESSELKPLTIINKNKNVTPIISRDKVINKLISE 2583
OY 755 GSN--LTSYGRTEAEFFAEAFRLMHSSTDHAERL---KVQKNAPKTFQFIND 801
Db 2584 YNNNDNVSYNAVAKLE--EDANNITIRDLDTSHMLNDLIQKN-----FKIIDD 2629

RESULT 5
026223 PRELIMINARY; PRT; 2269 AA.
AC Q26223;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE RHOPTRY PROTEIN.
OS Plasmodium berghei yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=5862;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YM;
```

```
RA MEDLINE-95021522; PubMed-7935623;
RA keen J., Sinha K., Brown K., Holder A.;
RT "A gene coding for a high-molecular mass rhoptry protein of Plasmodium
RT yoelii."
DR Mol. Biochem. Parasitol. 65:171-177(1994).
DR EMBL: L27838; AAA21304.1;
SO SEQUENCE 2269 AA; 265158 MW; F3D8CB103FB9A6E1 CRC64;

Query Match
Best Local Similarity 20.5%; Pred. No. 0.0083;
Matches 210; Conservative 155; Mismatches 307; Indels 350; Gaps 51;

OY 39 DVGHNVEKEKKNKDNKRKDE-----ENKTOEHLKEIMK---HIYKI---- 79
Db 1187 EVNSKIKNIETVYQHKKNYEIGIYVEKINETAKTNKNQIESTKELIKPTIOHIISFNAN 1246
OY 80 EVKG-----EFAVK-EAEKLLKVPDVLNEMKAIKGIYIYDGD 120
Db 1247 DLEGIDSDENLKYNTKNGNIEEFIKSYNLITLYLEVSKEST-TYNOIQNKRIDTQKE 1305
OY 121 ITKHIS-----LEALSED-----KKIKIDYIGKQALLHENVYAKEGEYEV- 161
Db 1306 LKNIENVNKAKSYLDYIKEREPRIYTHKKKINTV--NNEFKNE-YSKVNEGFDNISN 1362
OY 162 -----LYQSSDY--VENT-----EKALNYYEIGKI----- 187
Db 1363 SINTVKNSTDENSLLNLQTKEMYANIVNNTYYSKYEAEANIFRNIPKLANTLNIRIKN 1422
OY 188 -----LSRDI-----LSKINQPOKFLVNLNTIKNAS 214
Db 1423 SSGIDLSKDIKIALIYLSDKTDEDTLFIIPSPQKTEYITKISYSLIDILKSOELO 1482
OY 215 DSDGQ--DLF-----TNQLKEHPTDF-----SVEFLQNSNEVOEYFAK 252
Db 1483 KKEQOTLKLIPENRLEKVOATNELRGTLSDLYKKKEKILSEYKLLHKSNEMLKLSGN 1542
OY 253 AFAYIIEPOHADVQLYAPAFNYMDKFNQOEINLSLELK---DQKMSYKEWME---K 306
Db 1543 F-----QNYDTIL--ESSKY-DQVKEKSNYKQEKELGIDFVNTDMEKEFNNDIK 1590
OY 307 IKOHQMSDSLSEGRGLKLLQ-----IPIEPK---KDDIHSLSQEK- 349
Db 1591 VIEELNNDYS-SEENNINLOSRRKLLELTNAFNAEIKRIDDKIIEKNLDLIDKLETIRKN 1649
OY 350 -----ELKRIQIDSSDLSTE--EKEFLKQIDIRDSLESEKEL-----LN 391
Db 1650 CMLPHTTTLAETLKIKITDYKFIESAATKFSKEFLKYIG-OTSNSLNDIATLQKYDLH 1708
OY 392 RIQ-----VDSNPLSEKEKEFLKKLKDIOPIYDINORLODTGGLDPSINLQV 441
Db 1709 QINKYVTSKLSDATNDNNMLIEKEKE-----ATQAIKNLTLETTIDSNNIDA 1755
OY 442 KROYKRDIONI--DALHOSIGS--TLYNKIYVENMNI----- 476
Db 1756 NALNNKNTQWYFNSSELKHSIESIKQILKMKHVKRLNIGQINGKYDIDISQFONILQLO 1815
OY 477 -NNLTATLQADL--VDSNTDKINRGI-----FNEFKKNFKYSSISNYMIVDINE 523
Db 1816 ESELTANLN-DLKEIGQIKSDKKKKFLHALNETPIPNENTLKEIYHDIYVKKRQIDEEN 1874
OY 524 RPAIDNE-----RLKRIQLSPTTRAGYLENGKLLIQRNIGLEIKQVOIITKQSE 573
Db 1875 ITSEENENITLYIDITTKLKEKVOSILNFVTYENDSNIIKOHIODTNENNVSKIKESL 1934
OY 574 EYRIDAQVVPK-----SKIDPKIOEALNIOEW---NKALGL 609
Db 1935 TTIOSFOEILKINGIKAQFIDNNNNINNTISTISQDVNDVKKHISKDLTJENELIEI 1994
OY 610 PKYTKLI--TFNV--HNRYASINIVESAVALLNEMKNNTIO---SDLIKVTYVLYDGN 659
Db 1995 QKSLIEDIKNSTYEIRGANNNVNTIRNVYEQGTAKIQNNSKDEIDIDILNLYNKE-- 2052
```

QY	660	GRFVPTDITLPIAAQYTHODEITYQVNSKGLVPESSILLHGSKGY-----ELRND	713
Db	2053	-----SERKLPPTTKNNKNVYSIIISRI-NKYVNIILIESE---YGNNNNVSYVAKRLEED	2102
QY	714	SEGFIHERGHAVDDYAGYLLDPKNSQ---DLVYNSKKFIDIFK-----EEGSNLTASYGRT	764
Db	2103	ANSIILD-----LDRSQNIILDLIQNLKIIDLKKNKQOEIENRNLOTINRE	2150
QY	765	NE 766	
Db	2151	QE 2152	
RESULT	6		
Q9BJX9			
ID	Q9BJX9	PRELIMINARY;	PRT: 2747 AA.
AC	Q9BJX9:		
DT	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	235 KDA RHOPTRY PROTEIN (FRAGMENT).		
GN	IIA.2.		
OS	Plasmodium yoelli.		
OC	Eukaryota: Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=73239;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-YM:		
RA	Khan S.M., Jarra W., Peter P.R.;		
RT	"Distribution and Characterization of the 235 kDa Rhoptry Multigene		
RT	Family Within the Genomes of Virulent and Avirulent Lines of		
RT	Plasmodium yoelli."		
RL	Mol. Biochem. Parasitol. 0:0-0(2001).		
DR	EMBL: AF323443; AAKI5626.1; -.		
FT	NON_TER	1	
FT	NON_TER	1	
FT	NON_TER	1	
Q0	SEQUENCE	2747 AA; 321237 MW; 430508AC7B8824BD CRC64;	

Query Match	6.1%	Score 254.5;	DB 5;	Length 2747;
Best Local Similarity	18.8%	Pred. No. 0.011;		
Matches 196;	Conservative 168;	Mismatches 298;	Indels 379;	Gaps 44;

```

0Y      45 KKKKNNKKNRKRDKERK-----TOEBHLKE---IMKHVTKLEVGGEAVNKAEMAK 94
Db      1648 KKKKKIEOKKELDENVNSKIKNIENVTVBQHKNYEIGIVEKEINEIAKTNNKI--ESTAE 1705
0Y      95 LLEKPSDVULEMYKAIG-----CKIYLVDDITKH-----ISLEAL 130
Db      1706 LIKPITHIISSFNAXDLEGDIDSDNLGKXXHTEKXNNTIEEFKSYNLTTNYLETYSKESI 1765
0Y      131 SEDKKKINDYGKDALLHEHYVAKEGEPEVLVIOSSEDEVENTE-----KALANY 181
Db      1766 TYNQONKRIRDTQRLL-----KNIENVAKASYDIYIKENEFDRIYTHFFKKKLNV 1817
0Y      182 YEIGKILSRDLISKINOPYKFVLVDLNTIKNASDSGDODLTYNOLKE----HPDQS 235
Db      1818 MDNFK----NEKSXKNEGFDNISINTVKNSTDENSL-LNTLNQTEKYANIIVNNTYS 1872
0Y      236 VEPLEQNSNEVEQVAK-----AAAYIEPQHRYL--- 266
Db      1873 YKY-----EENIFRNIPKLANTLNIKNSSGIDLFRDIALSYLDSKTEDTLIFI 1926
0Y      267 -----OLVAPEAFNY-----MDKFNEOEINLSLE----- 290
Db      1927 PSPQKITTYYIKISDISYSILDLIDLKSOELOKKEOOTLKLFENRRLYEKVQATNELRGF 1986
0Y      291 ----ELKDORMLSYRE-----KWEEKIQ---HYOWMSDS 317
Db      1987 LSDDLKTRKEKILISEVKLLHHSNLENLKTSICNFQNTDITLESSEKYOVQVKEKSANNYOAEKPK 2046
0Y      318 L-----SEEGGLKKLKO-----IPREP 335

```

```

Db 2047 LGIDRNVADMEKFNNDIKVIEELENNDDSSSENNNTIOSKOKLWELTKKFAEIKKID 2106
Qy - 336 K---KDDIIHSISOEK-----ELKRIQIDSSDFLSTE---EKEFLKKLQIDI 378
Db 2107 KXIEKNDELIDKILFEPKKNOMLFTHHTLAFTLKIKITDYKFLIESATKFEKFLKYG-DT 2165
Qy 379 RDSISEEKEL-----LNRIQ-----VDSSNPISKEKEFLKLIKLDIOPYDINQ 423
Db 2166 SNSLNDIATLQTKYDLHQINKVYTSKSLDATTNNNNLIEKE-----ATQ 2212
Qy 424 RLDDPGGLDSSINDVAKQYKRQIQNI--DALLHQSIGS--TLNKKIYLENNMI---- 476
Db 2213 AAKNLTLETTIDSNNDANALHNKKIQWYVFENSELKSIESIKQLYKKMHVEKLLNIQOI 2272
Qy 477 -----NNLTATGLADL--VDSYDNKKIRGI-----FNEBK 505
Db 2273 NCKYDIDSKQFNNILOOSELTANLN-DLKEIGQKISKKKKFLHALNETPIPNFTLK 2331
Qy 506 KMFKYSISSNMIVDINERPALDNE-----RLKRIQLSPDTRAGYLENKLIIQ 555
Db 2332 ELYHDIIVKXKRQIDELENTISENENITLYDTFLKCKVQSILNFVYTEENDSIIINQ 2391
Qy 556 KNIGELKRVQIITKQSEKTYIRIDAKVYK-----SKIDTKIOE 594
Db 2392 HIQDTNENDVSKIESKLTKTIQSFQILKINGIKAKQAFYDNNNNINNTIISQDQVD 2451
Qy 595 AQLNTINQEM--NKALIGLPKYTKLI--FENNHRRAVSIVESAVYLLNEMKNNTOS--- 645
Db 2452 VAKHHSKDLTXENELIEICKSLEDIKNSYETIRSEQITVYVNTIRNYVEQQTNRKIONNSN 2511
Qy 646 -----DLKIKVNYLVADGNGRFVFTDITPLIAEQYTHODEIYEQVHSGKGLVPPESRIL 700
Db 2512 KKEIDDIQIKILNKNK-----SETKLPITGKNNTSIISRI--KNVILIESE--- 2560
Qy 701 LHGPSKGV-----ELRNDSEGFHEFGHAVDYAGYLLDKROS---DLVTSKKEFDIF 751
Db 2561 -YGNNNNNYSYNAKKLEEDANSIILD-----LDKSONILKLDIIOONLKLIIDDL 2607
Qy 752 K-----EEGSNLTSGRTNE 766
Db 2608 KKKKQEIENRNNLQTTINREO 2628

RESULT 7
Q9BK46 PRELIMINARY: PRT: 3130 AA.
Q9BK46:
DT 01-JUN-2001 (TREMBlrel.17, Created)
DT 01-JUN-2001 (TREMBlrel.17, last sequence update)
DT 01-JUN-2001 (TREMBlrel.17, last annotation update)
DE RETICULOCYTE BINDING PROTEIN 2 HOMOLOG A.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21101060; PubMed=11160005;
RA Triglia T., Thompson J., Caruana S.R., Delorenzi M., Speed T.,
RA Cowman A.F.;
RT "Identification of Proteins from Plasmodium falciparum That Are
RT Homologous to Reticulocyte Binding Proteins in Plasmodium vivax.";
RL Infect. Immun. 69:1084-1092(2001).
DR EMBL; AF312916; AAK19244.1; -.
SQ SEQUENCE 3130 AA; 370415 MW; 13D973DB89D802026 CRC64;

```

Query Match	6.0 %	Score 248.5	DB 5	Length 3130
Best Local Similarity	19.8 %	Pred. No. 0.023		
Matches 184	Conservative 146	Mismatches 274	Indels 325	Gaps 43
QY	44	VKEKEKNKDKRKRKEERKKTOEHLKELMKIVKLEVGEEAVKKEAEKL-----	95	
DB	1684	LKEKKNMEVEYKKMMMYVNVETE-KELIKRKKKNVEIMHIKIKETKKKKKEESNN	1742	

```

OY 96 -----LEKVPVLEMYKAIGKIIYVGDITKHSLE 128
DB 1243 KSLTTLMDSPFRSMFYNEIYNDYNINENPEKHQNLINELIYNGFNESYNIIMTKMEIINDN 1802
OY 129 ALSDEKKKIKDI-----YKG-----DALLHEHY---VYAKGEYEVLYIÖSSEDEVEMTEK 176
DB 1803 LDVYEIKEIKEVQATEYDKLKKVDEL--KNYLNINKEÖEGHRLI-----DYIK--EK 1851
OY 177 ALNAYVEIGKILSRDILS-----KINOPYOK-----FLDVLTNTIKNASDSGQDLF 223
DB 1852 IFNLYIKCSE--QONIIDDSYNTIVKKQYIKTIEDVKFLLDSINTLIEKKKSVANLEIC 1909
OY 224 TNO-----LKEH-----PTDPSVEFLÖNSNEVOEYAKAFAYIEPÖH----- 262
DB 1910 TNKEDIKNLKHVIKLANFSGIIVMSDNTIEITPENPLEDDLL--NLOLYFERKHEITS 1967
OY 263 -----RDVQL--VAPAFNYMDKNE----- 282
DB 1968 TLENDSDLELDHLSNDSDESIDNLKYNDITELHTYSTÖLLIKYDNIÖKLGDCNDLVKD 2027
OY 283 -----QEINLSLEELDKDQRMLSRYEKMEKIKOHYÖHMSDSLSEEGRGLKLOIPERPK- 337
DB 2028 CKELRELSTALYDKIQ--ITSVINRENDISNNIDIVSNKLENE-----IDAQYAFEKYE 2081
OY 338 -----DIIHSLSÖEKEFLKRIÖDSS-----DPLSTEKEFLKLOIDIRDSL 382
DB 2082 IFDNVEEYKTLDDTKMAYIVKKAELIKNVDTINKREDLDIYFNDLDEKSL-----TL 2135
OY 383 SEEEKELNRIÖVDSNPLSEKEKEFLKLKLDIQPYDINORLÖDTGGLDPSINDVR 442
DB 2136 SSNEMETIKTIÖV--NSYNSFS-----DINKNIN-----DID 2164
OY 443 KÖYKRDIONIDALLHÖ--SIGSTLYNKIYELENNINNLATFAGADL--VDSTDN----- 493
DB 2165 KEMKTLIPMDELINEGHNIDISLYNFI-----INIOIKIGNDIKNIREQENDRNIC 2217
OY 494 -----TKIRGIFNEFKKNFKYS--TSSNYMIVD-----INERPALDNER 531
DB 2218 FEYIÖNNYNFIKSDISIFNKYDHIKVDNYISNIDVANKHNSLSSEHVINATMIEN-I 2276
OY 532 LKWRIOQSPDTRAGYLENGK---LILÖRNIGLEIKDVÖIÖKSEKEYIRIDAKVYPRSKI 588
DB 2277 MTSIVEINEDTEMNSLEETÖKKLELEYENF-----KREKNINNNYKIVHEPKL 2325
OY 589 DTKIÖ---EAOININÖENKALGLPKYTKLITFENVHNRASYNIVESAYVLIENKKNIOS 645
DB 2326 -KEIENSLEYNTSISTJNFNKI-----NETÖNIDILKNEF--NNIKT 2363
OY 646 DLIKVYNYLVGDNGRVPFDITL-----PNIAÖYTHÖDEIYEÖVHSHK 689
DB 2364 KINDVK-----ELVHVDSTLTLESIOTFNNLYGDLMSNIÖDYKYEDINNVELKKV 2415
OY 690 GLVYPESSILLHPSKGVELRNDSEGFIFHEFGHADVAYG-----YL-LDKNÖSDLV 741
DB 2416 KLYENTNLL-----GRINTFIKELDKYÖDENNGIDIKYIETINKENNSYI 2460
OY 742 TNSKKEFLIDKEEGSNLTSGRTNEAEFF 770
DB 2461 IKLEKANNLNKENSKLÖNIKRNETEY 2489

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21101060; PubMed-11160005;
RA Triglia T., Thompson J., Caruana S.R., DeIorenzi M., Speed T.,
RA Cowman A.F.;
RT "Identification of Proteins from Plasmodium falciparum That Are
RL Homologous to Reticulocyte Binding Proteins in Plasmodium vivax.",
DR EMBL; AF312917; AAK19245.1;
SQ SEQUENCE 3254 AA; 382876 MW; 6F9CAFA5AA6167BA CRC64;

Query Match 6.0%; Score 248.5; DB 5; Length 3254;
Best Local Similarity 19.8%; Pred. No. 0.024;
Matches 184; Conservative 146; Mismatches 274; Indels 325; Gaps 43;

OY 44 VKEKKNKDEKRRKDERNKTOEHLKEMKHIYKIEVGEAVKKEAEKL----- 95
DB 1684 LKEÖNNAEYKKNEMVNYNETE--KEIHKKNYEIRIMEHIKKEETNEKKKFMESNN 1742
OY 96 -----LEKVPVLEMYKAIGKIIYVGDITKHSLE 128
DB 1743 KSLTTLMDSPFRSMFYNEIYNDYNINENPEKHQNLINELIYNGFNESYNIIMTKMEIINDN 1802
OY 129 ALSDEKKKIKDI-----YKG-----DALLHEHY---VYAKGEYEVLYIÖSSEDEVEMTEK 176
DB 1803 LDVYEIKEIKEVQATEYDKLKKVDEL--KNYLNINKEÖEGHRLI-----DYIK--EK 1851
OY 177 ALNAYVEIGKILSRDILS-----KINOPYOK-----FLDVLTNTIKNASDSGQDLF 223
DB 1852 IFNLYIKCSE--QONIIDDSYNTIVKKQYIKTIEDVKFLLDSINTLIEKKKSVANLEIC 1909
OY 224 TNO-----LKEH-----PTDPSVEFLÖNSNEVOEYAKAFAYIEPÖH----- 262
DB 1910 TNKEDIKNLKHVIKLANFSGIIVMSDNTIEITPENPLEDDLL--NLOLYFERKHEITS 1967
OY 263 -----RDVQL--VAPAFNYMDKNE----- 282
DB 1968 TLENDSDLELDHLSNDSDESIDNLKYNDITELHTYSTÖLLIKYDNIÖKLGDCNDLVKD 2027
OY 283 -----QEINLSLEELDKDQRMLSRYEKMEKIKOHYÖHMSDSLSEEGRGLKLOIPERPK- 337
DB 2028 CKELRELSTALYDKIQ--ITSVINRENDISNNIDIVSNKLENE-----IDAQYAFEKYE 2081
OY 338 -----DIIHSLSÖEKEFLKRIÖDSS-----DPLSTEKEFLKLOIDIRDSL 382
DB 2082 IFDNVEEYKTLDDTKMAYIVKKAELIKNVDTINKREDLDIYFNDLDEKSL-----TL 2135
OY 383 SEEEKELNRIÖVDSNPLSEKEKEFLKLKLDIQPYDINORLÖDTGGLDPSINDVR 442
DB 2136 SSNEMETIKTIÖV--NSYNSFS-----DINKNIN-----DID 2164
OY 443 KÖYKRDIONIDALLHÖ--SIGSTLYNKIYELENNINNLATFAGADL--VDSTDN----- 493
DB 2165 KEMKTLIPMDELINEGHNIDISLYNFI-----INIOIKIGNDIKNIREQENDRNIC 2217
OY 494 -----TKIRGIFNEFKKNFKYS--TSSNYMIVD-----INERPALDNER 531
DB 2218 FEYIÖNNYNFIKSDISIFNKYDHIKVDNYISNIDVANKHNSLSSEHVINATMIEN-I 2276
OY 532 LKWRIOQSPDTRAGYLENGK---LILÖRNIGLEIKDVÖIÖKSEKEYIRIDAKVYPRSKI 588
DB 2277 MTSIVEINEDTEMNSLEETÖKKLELEYENF-----KREKNINNNYKIVHEPKL 2325
OY 589 DTKIÖ---EAOININÖENKALGLPKYTKLITFENVHNRASYNIVESAYVLIENKKNIOS 645
DB 2326 -KEIENSLEYNTSISTJNFNKI-----NETÖNIDILKNEF--NNIKT 2363
OY 646 DLIKVYNYLVGDNGRVPFDITL-----PNIAÖYTHÖDEIYEÖVHSHK 689
DB 2364 KINDVK-----ELVHVDSTLTLESIOTFNNLYGDLMSNIÖDYKYEDINNVELKKV 2415
OY 690 GLVYPESSILLHPSKGVELRNDSEGFIFHEFGHADVAYG-----YL-LDKNÖSDLV 741

```



```
Db 2416 KLIENITNL-----GRNTEFIKELDKQDENNGIDKIEIENKENSXI 2460
Qy 742 TNSKFFIDFKEGSNLISYGTNEAEFF 770
Db 2461 IKLEKANNLKENSKLONIKRNETELY 2489

RESULT 9
ID 025662 PRELIMINARY; PRT: 1939 AA.
AC Q25662;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)
DE REPEAT ORGANELLAR PROTEIN.
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5825;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=96V;
RA Weiner E.B., Taylor W.R., Holder A.A.;
RL Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.
DR EMBL: U43145; AAC63403.1;
SQ SEQUENCE 1939 AA; 229001 MW; B36E462001C6F22F CRC64;
```

Query Match 5.9%; Score 246.5; DB 5; Length 1939;
Best Local Similarity 19.6%; Pred. No. 0.016;
Matches 198; Conservative 174; Mismatches 302; Indels 335; Gaps 50;

```
Qy 36 GHGV--GMHYKKEKKNDENKRRKDEENKQOEHLKEIM-----KH---IKYIEYKG-- 83
Db 898 GHREYVAGLEEKHKKEVVALEEKHEKLEAKKEGH--KEYMELGKHKKEVVAAGLAKHNL 956
Qy 84 EEAVKKEAEKLEKVPDVL-----EMYKAIGKIIYVDGDTK-----HISLEALSEDKK 135
Db 957 EEGHKEMVAE--LEKRNADLVAVLEQHKH-----EIKIGEEKKEVVAAGIEEYK 1004
Qy 136 KIKDIYGDALLHEHYVAKEGEYVAVLIQSSSEYVENTKALNYYEIGKILSRDILSK 195
Db 1005 KVEAL---KLAEEH---KD-----VYTKLGEQHKKEELAKLEDGKHEVVAEVEKKNSL 1051
Qy 196 INOPQKFLDVNTIKNASDSQODLFTNOLKEHPTDESVFLDONSVEQVAVAKAFA 255
Db 1052 LAMLEENHKNEKIKLEKHKESASDLVEKLYOKDEEVKSNKKIEELTVNIDL----- 1105
Qy 256 YVIEFQHRDVLQYAPFAFNVDKFN--QETN--LSLEELKD----- 294
Db 1106 -----NDSIMCYKKQJLLEVEKKNENVEINLKLIYQNEKMDKDKILEKENIKKL 1158
Qy 295 QRMISRYEKWEKIKOHYH-----WSDSLSE-----EGRGILKKIQPIE 334
Db 1159 NKKLSNYVFETKENTYKNSKENVVNEKERIIVDSVCKENISESDVEGGMILKMTLSL 1218
Qy 335 PKKDDIHSLSOEKELLKRIQIOSSDPLSTPEKEFLKLOI----- 376
Db 1219 -KKEKNITSINDKNK-----SSELVDYTKSAVINKIEYKKEIEDNGKNIEDLKNK 1269
Qy 377 -----DIRDSISEEKELLNRIQVDSNPULSEKEK-----EFLK-----KTLK 414
Db 1270 ILDSLNLINLENKKNVLTDENNNLKKEIET--KDKLKEKKEKNEMTEILLNLNDIITKLK 1328
Qy 415 DIQPYDINQRLQDTGGGLDPSINL-----DVAKQYKRDIONIDALLHQSIG--STLYNK 467
Db 1329 EISEW-----KDEEEKLTKENIKLKNDIEQINKEYKIKEERILMKFENINEVSLKNQ 1382
Qy 468 IYL-----YE-----NANI--NMLTATIGADLVDSYDN--TKINRIFNE- 503
Db 1383 IEIEKMKLEELKNTELLLAERETNMSISDNDKNIVENNILLKQNNKNKNVEVK 1442
Qy 504 -----FKKNFYSSISNMIYDINERPALDNERLKMRIQLSPDRAGYLGNGKLLIQRI 558
```

```
Db 1443 GDDINCEKNNDQAKETISYLKDEIKTISMLYGEPL-----NRKNSYDEKAK-----NL 1489
Qy 559 GLEIKDVQI-----IKQSEK-----EYIRID 579
Db 1490 TNELKELKIRNKGGEAEIAELNKKIKERKNKSQVKNDESSSNITTKDGDKTPEVSD 1549
Qy 580 AKVVP--KSKIDTKIOEADLNINOENKALGLPKYIKLITFNVHNKNASNIYESAVLIIN 637
Db 1550 DRIOKDKANLVLKLEKP-----DLMDNINSLEK-----ENFRVMSIYKE----- 1590
Qy 638 EWMKNIOSLIKKVTNYLYVDGNGR-----FVFTDI--TLPNIAEOYTHODEIYEQVH 687
Db 1591 --NRKNVQNKIYGIYIFPKKCEKELKNDLVYICLVAKDLISILFLDNVNLFEKIDKIL 1648
Qy 688 SKGLVVP--BSRSTLLHGSPGVELRN--DSEGFI--HEFGAV-----D 726
Db 1649 WKQWYIPTEIRILFLRYFSLDKLRNYKCVNEEYVNNERYEYSMALFQYLETASNLK 1708
Qy 727 DYAGYLLDKNOSDLVYNSK-----KFIDI-----FKREGSNLT 759
Db 1709 EMIIYVLEKAEKDCSCENSNDFKPKITDILNFSKDSIRLKTIAQLRKELENFEREAKNIL 1768
Qy 760 SYGRTNEAEFFFAFRLMHSTDHAERLKYOKNAPKTFQF--INDQIKETII 807
Db 1769 NYDYOIILKKHYECLR-----KLKIVKMARLEDRNYVSSKFSI 1808
```

Query Match 5.9%; Score 245.5; DB 12; Length 1127;
Best Local Similarity 20.6%; Pred. No. 0.0088;
Matches 167; Conservative 126; Mismatches 263; Indels 253; Gaps 37;

```
Qy 09YVT6 PRELIMINARY; PRT: 1127 AA.
ID 09YVT6;
AC Q9YVT6;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE ORF MSV156 HYPOTHETICAL PROTEIN.
GN MSV156.
OS Melanoplus sanguinipes entomopoxvirus (MSEPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OX Entomopoxvirus B.
RN [1]
NCBI_TaxID=83191;
RP SEQUENCE FROM N.A.
RC STRAIN=TUCSON.
RX MEDLINE=99102612; PubMed=9847359;
RA Alfonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;
RT "The genome of Melanoplus sanguinipes entomopoxvirus";
RL J. Virol. 73:533-552(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TUCSON.
RA Alfonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;
RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF063866; AAC97677.1;
SQ SEQUENCE 1127 AA; 134265 MW; F185DA1D5A3FETD1 CRC64;
```

Query Match 5.9%; Score 245.5; DB 12; Length 1127;
Best Local Similarity 20.6%; Pred. No. 0.0088;
Matches 167; Conservative 126; Mismatches 263; Indels 253; Gaps 37;

```
Qy 60 ENRKTQOE--HLKELMKIIVIEYKGEAVVAKKEAEKLEKVPSPVLEMYKAIGKIT- 115
Db 24 ENNKVSLDINSLSYELTNI-----KFSDKITMEIILNKNYKVIYEKIFY 66
Qy 116 -----YVDGDTTKHISLEALSEDKKIKD-----TYGDALLHEHYVAKGEYEPV 161
Db 67 MHNQKINDYNIILQYLIEFYNNIEINKCTIKENKFPCKNPPLY-----NITTKKKIITYD 118
Qy 162 LVIQSSSEYVENTKALNYYE-----ICKILSRDILSKINOPY-----QKFLDVL 207
Db 119 L-----DYEEKKDKELVINIQKNAVDKI--NDIKNNVNNHISDNETIITGKETLIDIL 170
```


DE PFG377.
 OS Plasmodium falciiparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=96360472; PubMed=8719156;
 RX Handman E., Osborn A.H., Symons F., van Driel R., Cappai R.;
 RT "COS cell expression cloning of Pfg377, a Plasmodium falciiparum
 RT gametocyte antigen associated with osmophilic bodies.";
 RL Mol. Biochem. Parasitol. 74:143-156(1995).
 DR EMBL; L04161; AAC37257.1;
 DR InterPro; IPR001064; Cystealin.
 DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
 SO SEQUENCE 3119 AA; 377358 MW; 519B99D25BDEFCC CRC64;

Query Match 5.8%; Score 239.5; DB 5; Length 3119;
 Best Local Similarity 19.9%; Pred. No. 0.055;
 Matches 182; Conservative 153; Mismatches 313; Indels 265; Gaps 46;

OY 27 FIPVAGAGHGD-----VGMHVKEKKNKDNKDKDERNKTOEHLKEIKHIVKIE 80
 DB 964 FTTISGEENHNRRELKKIEANLKEEMKRFNEQOEERKKABE--DENMETIQKH 1021
 OY 81 VKGEAEVKEAEKLEKVPDYLEMYKAIGKITYVDG---DITKHISLEALSD--K 134
 DB 1022 METSKIEKEEVDYQDEPD-----KQYAIHQOELEFLRTDSEGESDVKR 1071
 OY 135 KKIK-----DIYKDALHEHYVYAEYEPVLVIOSSDYV-ENTERALNVEYET- 184
 DB 1072 DKVKPPRDPSPDFYNNALISFHEKMEELYN--TSISSLYVKEINRKFDDYKELK 1129
 OY 185 -----GKLSDILSKINQ-----PYQ-----KPLDVLNTINADSDG 218
 DB 1130 SKTYPEKEDLTSGOTNCKNKLFOKLETKDEYOKNIOGYKNKVIDLDDIQ--KKANG 1187
 OY 219 ODLEFTNOLKEHPTDFSVFLEONSNEVOEFAKAF----- 254
 DB 1188 KTIILIONLIEIKIDYKGVNRLSDRKFKNFRKVLGKRKKMLLEDFFRAQFKAIRFIK 1247
 OY 255 --AYIEPQHRYLQ-LYAPEAFNYMDKFNEDFINLSLEBKQRMASR--YEKKEKIK 308
 DB 1248 LTTTVEEETVLYEDIVYMEKKYKEEYS-----KMRRIISSMLDEYVAKQIK 1296
 OY 309 OHYOHMSDSLSEGRGLKKIQIPIEPKKDDIHLSOSEKELKIQIDSSPFLSTEEK 368
 DB 1297 EHY-HKVDITISEH-----KFOEIROHMRDKIENTIHELKEMYVQIQIDLTNYV- 1344
 OY 369 EFLKLIQIDIRDSLSEKELNRIQVDSNP-----LSEKEEFLKKLQIDPYDIN- 422
 DB 1345 -----HOLENHISLLOALQONKNIPRHNLVLEKKELEITRRKKKN-KP-DISIS 1391
 OY 423 -----QRLQDTGGLIDSPSINLVRKQYKRDIONIDALLHOSIGSTLYNKIYLYENMN- 475
 DB 1392 SHATDEQOVSDF--LIRGAHNHGDIIKGEDND---EVLLIEQIQSL--KTMGONQ 1442
 OY 476 -----INNLATLGAADVDS-----TDWTKINRGJFNEKKKFKSISINVIIVINER 524
 DB 1443 VGSILEKLNLS-----DOYQLQDKLNVVEDIYKML-RNFKHYIEKLHESKINRE 1493
 OY 525 PALDNERLKWRIQLSPDTRAGYLENGKLLORNIIGLEIKDYQ-----IIKOS 571
 DB 1494 KFI-----TVVDVLSNYSFLELYVVKFLHLHDFQMSFEKDELEKHLYLELE 1539
 OY 572 EKEYIRIDAKVVPKSKIDTKIQEAO--LINOENMKALGLPKYTKLITFNHNRVANSI 628
 DB 1540 RKKYITLLEIOT--RDPLSTNIOQEGDHIINNMMNN-----NVARNMLKKQY 1583
 OY 629 VESAVYLMEKKNNI-QSOLIKYV-----NYLVGNGRVEYFDITLPNTAEQYTHODEIT 663
 DB 1584 LADDLLEISKLGHILEVDIKKNIALQOINYLNN-----TNETVDPVIRDLMPAPRIY 1637

OY 684 ---EOVHSGKLYVESKSTILLHGPSKGVLELRNDSGEFHFEFGHADVADYACLLDKNOSDL 740
 DB 1638 PVSEDIYDVTWRDNTAV-----INNLRHRHVMFDDKIQYDDHLL-----F 1681
 OY 741 VTNSKKFIDIFKE-----EGSNLTSGRTNE-----AEFFAEARLHNSIDHAERLQOKN 791
 DB 1682 VYNIKEL--TYKENLADKERNANYEYEMNRFYLHLEEFF--YILKH---YVELRKIQDL 1733
 OY 792 APRTQFINDQIK 804
 DB 1734 GEVAIRPSFEENIK 1746

RESULT 13

O9EMP3

ID O9EMP3

PRELIMINARY;

PRT; 1238 AA.

AC O9EMP3;

01-MAR-2001 (TREMBLrel, 16, Created)

DT 01-MAR-2001 (TREMBLrel, 16, last sequence update)

DT 01-MAR-2001 (TREMBLrel, 16, last annotation update)

DE AMV156.

GN AMV156.

OS Amsacta moorei entomopoxvirus (AmePV).

OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;

OC Entomopoxvirus B.

OX NCBI_TaxID=28321;

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20396580; PubMed=10936094;
 RA Bawden A.L., Glassberg K.J., Diggins J., Shaw R., Farmerie W.,
 RA Moyer R.W.;
 RT "Complete Genomic Sequence of the Amsacta moorei Entomopoxvirus:
 RT Analysis and Comparison with other Poxviruses.";
 RL Virology 274:120-139(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Bawden A.L., Glassberg K.J., Diggins J., Shaw R., Farmerie W.,
 RA Moyer R.W.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF250284; AAG02862.1; -;
 SO SEQUENCE 1238 AA; 147096 MW; 9BFB0A39DB6E35B CRC64;

Query Match 5.8%; Score 238.5; DB 12; Length 1238;
 Best Local Similarity 21.0%; Pred. No. 0.019;
 Matches 189; Conservative 152; Mismatches 276; Indels 281; Gaps 49;

OY 56 RKDEERNKTOEHLKEIKHIVKIEYGEBAVKEAEKLEKVPDYLEMYKAIGKI- 114
 DB 158 RSDQEFNINIKNYIHDIKN-----DNNIIDQLISRPIIDIIKKEFIDKLT 205
 OY 115 -----YIVDGDITKHIS-----LEALSDEKKIKYIDGKALLHEHYV---AKEG 157
 DB 206 NYIKNYISDKLVIINISKNPLFNINIKIYINENIKINNIINIK-IVNDDFLYFINNYNKN 264
 OY 158 YEPVLVIOSSDEVNEMKALNVEYEIGKILSD-----IISK-----INQPYOKF 203
 DB 265 YDFINLKNNEEDTIFLKNITINDY-INNLTSSDEYKNIITDMETISKNIDITMDYVEKN 323
 OY 204 LDVNLNTIKNASDSG--QDLFTNQ--LKEHPDVSVEFLEONS-----NEVG 247
 DB 324 SDVIKRLKQKISDDEVYLSNLTSENENFLKKIKIISSENIDISLNSLNINENKFKILK 363
 OY 248 EVFAKAFAYIEPQHRDYLQ-----YAPEAFNYMDKN--EOEINLSLEELK--- 293
 DB 384 EIINKFEL-----KGNIDIMLVNPNFYDKQKSIDEKIDIDINIKODEIYTHLSISDLVLI 438
 OY 294 -----DORMLSRYEKWEKIKOHYOHMSDSLSEGRGLKYL 329
 DB 439 SPQNLFTYNDIDTMIESTYDKQILNLVQETEKLHDTYKQVANSRPRQ---LSDV 494
 OY 330 QIPIEPRKDD-----IHSLSQEB-----KELKRIQIDSSPFLSTEEKEFLKIQ 375

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:54:01 ; Search time 132.69 seconds
(without alignments)
857.637 Million cell updates/sec

Title: US-09-747-521-2_COPY_1_778

Perfect score: 3987

Sequence: 1 MNKKKFFIKVINSCLVTAI.....TSYGRTEAEFAFAFRMLH 778

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL_17:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	273.5	6.9	2752	5	Q9BJYO plasmodium
2	268	6.7	1946	5	Q97291 plasmodium
3	255	6.4	2269	5	Q26223 plasmodium
4	254.5	6.4	2747	5	Q9BJX9 plasmodium
5	253	6.3	1979	5	Q96133 plasmodium
6	248.5	6.2	2771	5	Q26216 plasmodium
7	248.5	6.2	3130	5	Q9BK46 plasmodium
8	248.5	6.2	3254	5	Q9BK45 plasmodium
9	245.5	6.2	1387	5	Q9G276 plasmodium
10	239	6.0	1127	12	Q9YV76 plasmodium
11	237	5.9	1238	12	Q9EMP3 plasmodium
12	235	5.9	1048	5	Q26023 plasmodium
13	235	5.9	3119	5	Q25857 plasmodium
14	233.5	5.9	1939	5	Q25662 plasmodium
15	228.5	5.7	1786	5	Q9U0P0 plasmodium
16	225.5	5.7	1931	5	Q9NCF9 plasmodium
17	224.5	5.6	806	2	Q9L8P7 mycoplasma
18	224.5	5.6	1558	5	Q96275 plasmodium
19	221.5	5.6	3724	5	Q77320 plasmodium

20	220	5.5	2166	2	Q51465 borrelia bu
21	218.5	5.5	1931	5	Q9VKH9 drosophila
22	218	5.5	980	5	Q96246 plasmodium
23	217	5.4	1088	5	Q00905 oxytricha f
24	216.5	5.4	2867	5	Q9K2M5 plasmodium
25	214.5	5.4	1313	10	Q9XIP6 plasmodium
26	213	5.3	1516	4	Q96154 plasmodium
27	211.5	5.3	1206	4	Q9NWC1 homo sapien
28	209.5	5.3	976	2	Q54222 staphylococ
29	209.5	5.3	1005	1	Q58718 methanococ
30	209.5	5.3	1128	10	Q9SAP6 gsaite arabidops
31	209.5	5.3	1780	4	Q9UFR5 homo sapien
32	208.5	5.2	999	5	Q9U5A3 plasmodium
33	207.5	5.2	683	2	Q50281 mycoplasma
34	207	5.2	849	2	Q9AHR4 borrelia bu
35	206.5	5.2	1650	5	Q77328 plasmodium
36	206.5	5.2	1734	5	Q9U0L7 plasmodium
37	205	5.1	1065	2	Q9AHR8 borrelia bu
38	204	5.1	1065	2	Q9AHR7 borrelia bu
39	203.5	5.1	2139	5	Q07569 entamoeba h
40	203	5.1	849	2	Q9AHL2 borrelia bu
41	203	5.1	3259	4	Q14789 homo sapien
42	202.5	5.1	1070	10	Q9AW67 guillardia
43	201.5	5.1	841	5	Q23037 arabidopsis
44	201.5	5.1	1909	5	Q25893 plasmodium
45	201.5	5.1	1938	13	Q91BD7 seriola dum

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	2752 AA.
Q9BJYO				
AC	Q9BJYO			
DT	01-JUN-2001 (TREMBlrel. 17, Created)			
DT	01-JUN-2001 (TREMBlrel. 17, last sequence update)			
DT	01-JUN-2001 (TREMBlrel. 17, last annotation update)			
DE	235 KDA RHOPTRY PROTEIN (FRAGMENT).			
GN	IIIA.1.			
OS	Plasmodium yoelii yoelii.			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=73239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=YM;			
RA	Khan S.M., Jarra W., Peter P.R.;			
RT	"Distribution and Characterization of the 235 Kda Rhoptry Multigene Family within the Genomes of Virulent and Avirulent lines of Plasmodium yoelii."			
RL	Mol. Biochem. Parasitol. 0:0-0(2001).			
DR	EMBL: AF323442; AAK15625.1; -.			
FT	NON_TER	1		
FT	NON_TER	2752		
SEQUENCE	2752 AA; 322606 MW; E834E3753168AF76 CRC64;			

Query Match 6.9%; Score 273.5; DB 5; Length 2752;
Best Local Similarity 19.6%; Pred. No. 0.0017;
Matches 191; Conservative 158; Mismatches 283; Indels 343; Gaps 44;

QY	58	DERNKTOGEHKLKMKHKVIEVGEAV-----KKKAELKLEKYPVLEVKAI	110
DB	1736	NKEMNTTFEPMESINLIVDYLETYSKESITYGELKNNKISTOKELKSEIENVK-----	1790
QY	111	GSKIYVDGDIKHSLEALSEDKKIKIDYGRDALHHEVYAKGEYV-----	161
DB	1791	-AKSTL--DITFNEPDRIVTHPKKKINDVNDK---FTNEYSKVKNGFENINSINNVAK	1844
QY	162	-----LVQSSSEDIYVENTK-----ALNYYEIGKILS-----	189
DB	1845	SYDENLLSLNLOTREMYANIVSKRYYSKYEAENIFINISKLANSLNIOIONSSGIDLH	1904

```
QY 190 -----RDIIS-----KINOPYOKFLDYLTNTIKN-----ASDSGO 219
D 1905 KNINIAIVSYLNSOKEDMSIFSPQKTSYKISDSYFTLLDIFKSYXELORKEGOAL 1964
OY 220 DLFF-----TNOLKEHPTDF-----SVEFEONSNEVOEFKAFAYIE 259
D 1965 NLIFERKLHDKVOATNELKDTLSDLKNKEQLINKVXLLHKSNELNKSCS----- 2018
OY 260 POHRDVLQVAPAFNPMKFNDOEINLSLELK-----DORMLSRYEKWEKIKO 309
D 2019 -QNYDTIL-----ESSKY-DKIKESNNYEOEKNLGIDFVTAAMEKFNNDIDIELEN 2072
OY 310 HYOHMD-----SISEGRGLL-----KKLQ-----IPIPKAD-----IHSLSOEK 349
D 2073 NYKH-SDKDNYNPFSENNNNLQSKKELSMNPFNAEIKKXEDKIEKNKLKLETRK 2131
OY 350 ELK-----KRIQI-DSSDFLSTE--EKEFKLKLDIDIRDSSEBEKL-----L 390
D 2132 ECHFEFTYTLVLEALRIKIDYSKFITSATKFSKFLKYID-DTNSLNDINDINTLQTKYDL 2190
OY 391 NRIO-----VDSNPLSEKEKEFLKLLKLDIOPYDINORLODPTGLDPSIIND 440
D 2191 NOJKAHVTSMEADATNDNNLLEKEKATKTINNLELF-----TDSNNIDAD 2239
OY 441 VRQOYRDIONIDALLHOSIGS--TYNKTYLYENMNINNLATLGCADLVDSIDN----- 493
D 2240 VLNHNKTOMLYFSELSKIDSIKOLYKMKHVKFLINIGHINKRY-FDISKEFDNLOLO 2298
OY 494 -TKINGIFNEFKK-NFKYSSSNMYIVDNERP-----ALDNERLKWRIQLSPDTR 543
D 2299 ESLEKLEN-NDLKKIGOKISDKKNKFLNELSELPINPENTLKEIYHIVYESQIDEIKN 2357
OY 544 AGYLENGKLILQNIQLEI-----FVYTYENDSNVYIKOHIONDENDVSKIKONL 2417
D 2358 ITNENDNITLYMDITLTKMKVYESILNFYTYENDSNVYIKOHIONDENDVSKIKONL 2417
OY 574 EYRIADAKVPR-----SKIDTKIQEOQLMINOEM--KNALGLPR- 611
D 2418 KTESQOKILINKNETKAQOYDNNNNINNVISTISODVIDVKHISDLYXENELIEIQKS 2477
OY 612 --YTKLITFNVRHRYASNTIVESAYLLNEMKNIIQS-----DLIKKTYNVLVDGNGR 661
D 2478 LEVTKSTDIRSEQITKYVNPIDHYVEOOTKKIIONPNKDELDLOELVNTNKE----- 2533
OY 662 FVETDITLPR-----IABQYTHODEIYEOVHSGLYVPESRSIL 700
D 2534 --SEKLPITLINNKDNYPIIIRIDKVINLIKSEYNNNDNVSNVAKK----- 2579
OY 701 LHGPRSGVELRNDSEGFIEFGHAVDYAGYLLDKNQ---SDLYTNSKKFIIDFK----- 752
D 2580 -----LEEDANSTIRD-----LDTSONMLNDLLOKMLTIIDDKNNKQE 2618
OY 753 -SEGSNLTSGRTNE 766
D 2619 IENRXNLQITINRQE 2633

RESULT 2
ID 097291 PRELIMINARY: PRT: 1946 AA.
AC 097291:
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE HYPOHETICAL 231.8 KDA PROTEIN.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota: Alveolata: Apicomplexa: Haemosporida: Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=99376085; PubMed=10448855;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
```

```
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagsels K., Jassal B., Kyes S., McLean J., Mole S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajadream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.,
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT falciparum .";
RL Nature 400:532-538 (1999).
DR EMBL: AL034559; CAB39037.2; -.
DR InterPro: IPR002048; EF-hand.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_2.
KW Hypothetical protein.
SQ
SEQUENCE 1946 AA; 231792 MW; 59AC248AB9808E34 CRC64;

Query Match 6.7%; Score 268; DB 5; Length 1946;
Best Local Similarity 21.5%; Pred. No. 0.0019; Indels 208; Gaps 43;
Matches 183; Conservative 153; Mismatches 306;

OY 47 KKKNKDEKKRDEERNKQ-----EHLKEIMKHIVKIEVGEEAVK 89
D 1051 RDKNNHNNNNKNNKNNNNNNYVYQHNNNLSHITVLEKKKALKOKYVLENNILVOKK 1110
OY 90 EAA-----EKLEKVDSDVLEMYTKAIGKTY---IVDGIYHISLEALSEDKKIKDT 140
D 1111 EMFECNKKKKYKKKISLNEYKKLDNIITDFNKLKENCITKEKLLAKMEDITKYINEQ 1170
OY 141 YGRDAL-----LHEHYAAKEGEPVLYIOSSEDPYENFEKALNVYELKILSRDLK 195
D 1171 FSLSKIOFENKMDYVFLKKKSEIYML--KELKEKKITLYNDNLKQYKKDDOI 1227
OY 196 INOPYOKFLDYLTNTIKN-----ASDSGOQLLFTNOLKEHPTDSVEFLSEONSNEVOEFA 251
D 1228 LKENIEKIDIRKKLTKQEEIISOKDRQIETLENNK-----IGKERINKFNEIOQLQY 1282
OY 252 KAFAYIEPQ-----HRDVLQVAPAFNPMKFNDEIYL--SLE-LNDORMLS----- 259
D 1283 K-INIHEKETETKTNDIEKERNTKLSNKFIDINKNNNLHNKIFETLNEKEISTENV 1341
OY 300 -----RYEKWEKIKOHYOHMSDSLSEGRGLKLDIPIEPK 337
D 1342 KLETKNTLYIENKLSNDLKNLTKKEKIQENY---KNINKEHKIIELOSYPEIKE 1397
OY 338 DDI-----IHSLSOEKELLRIQIDSSDFLSTEKEFLKLDIQRDS--LSEEE 366
D 1398 KHKKEITQIEQOIHNDKH---IAQIOIEKNPF---ESYLKKNENKMSNILEEKY 1449
OY 387 KEL-----LARIQY-----DSSNPLSEKEKEFLKLLKLDIOPYDINORLODPTGLID 433
D 1450 KELSTYEIDKNINKIKIEDLEKKEENILTKNEE-INNLKEEYKM--VOOHLDETNYLYE 1506
OY 434 SPISIND-VRQOYRDIONIDALLHOSIGSTLYNKIYLYENMN-----INNLATLGCADL 487
D 1507 KOKLADITITKEKNNTINECDKIKNN--KLLNKK--LKENONNYETILLNTI----- 1554
OY 488 VDSNTDNKINGIFNEKKKNFKYSISS-----NY-MIVDINERPALDNERLKWRIQLS 539
D 1555 -----KKENOQIIEREKKNFTQKVESLEHAFKOSYQNLKDQENMLQOQIKOLK-NVNO 1607
OY 540 PDRAGYLENGKLILQNIQLEIKDVQIITKOSKEEYRIADAKVPRKIDTKIQEOQLNI 599
D 1608 IKTNKSLKLVNELL-----IKETKNYSQOKKEFKIK-GLKINKQAYIKLKNENQOLKI 1659
OY 600 NO-EMNKALGLPYTKLITFNVRHRYASNTIVESAYLLNEMKN-NIOSDLIKKVTNVLVD 657
D 1660 NAFEYIKKDVQDN---VTLNVHNN-----ILNOKKLPVQIDILKS---QVD 1701
OY 658 GNGREYFTDITLPNIAEQYTHODEIYEOVHSGLYVPESRSILLHGSKVGELELN----- 712
D 1702 OKONII-----NNMKEO-----IEDVNHKIASINKKEEL---NTYIKIKNKTTED 1744
OY 713 ---DSEGFIEFGHAVDYAGYLLDKNQSDLYTNSKKFIIDFEEGSNL--TSGRTNEAE 768
```


Db 1745 VNLSEVKSSELSNKKDEVKRRKTRIEIKQKE--REYKKLLDDYKIEKKNLVTKYK--EID 1800
QY 769 PFAEAFRLMH 778
Db 1801 STMYKFEPAH 1810

RESULT 3
Q26223
ID Q26223 PRELIMINARY; PRT: 2269 AA.
AC Q26223;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE RHOPTRY PROTEIN
OS Plasmodium berghei yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5862;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-YM:
RX MEDLINE=95021522; PubMed-7935623;
RA Keen J., Sima K., Brown K., Holder A.;
RT "A gene coding for a high-molecular mass rhoptry protein of Plasmodium
RT yoelii.";
RL Mol. Biochem. Parasitol. 65:171-177(1994).
DR EMBL: L27838; AAA21304.1; -
SQ SEQUENCE 2269 AA; 265158 MW; F3D8CB103FB9A6E1 CRC64;

Query Match 6.4%; Score 255; DB 5; Length 2269;
Best Local Similarity 20.5%; Pred. No. 0.0081;
Matches 210; Conservative 155; Mismatches 307; Indels 350; Gaps 51;

QY 39 DVGHHVKEKKNKKDKRDE-----ERNKTOEHLKEITK---HIVKI---- 79
Db 1187 EVNSKIKNIEMTVNQHKMYEIGIYEKINEIAKTNKNOQESRKELIKPTIQHIISSFNAN 1246
QY 80 EVKG-----EEAVKK-EAAEKILEKYPVSLVLEMYKAIIGKTIYVDCD 120
Db 1247 DLEGSDENLGKYTEMGNIYEERIKSYNLITNLETYSKESI--TYNOIQMKRIDTOKE 1305
QY 121 ITKHIS-----LEALSED-----KKIKIDYKDALLHEHYVYAKEGEYEV-- 161
Db 1306 LKKNENNNKAKSYLDYIKENEFDRIVTHFKKKLNTV--NDPKFNE-YKVNDEGEDNIN 1362
QY 162 -----LVYQSEDEY---VENT-----EKALNVYEYEGKI----- 187
Db 1363 STNTVKNSTDENSLNLTINQTEKEMANIYNNNTYSSYKAEANIIFRIPKLANTLNLIKIN 1422
QY 188 -----LSRDI-----LSKNOPYQKRLDVLNTIKMAS 214
Db 1423 SSGIDLSKDIDKIALISYDSKTEDTLIFSPQKRTETVTKTISDSYSLDLILKKSQELQ 1482
QY 215 DSDGQ--DLTF-----TNOLKEHPDVF-----SVEFLQNSNEVOEFAK 252
Db 1483 KKEGOTLKLIFENRRLYEKVQATNELLRGTLSDLKAYKKEKIISEVALLHKSSEINKLSGN 1542
QY 253 APAYITEPOHRDVLQVAPFAFNVDKNEQFINLSLEELK--DQRLMSRYEKME--K 306
Db 1543 F-----QNDYTLI-----ESSKY-DQVKEKSNNYKQEKELGIDFNVTYDMEKEFNNDIK 1590
QY 307 IKOHQYHMSDSISEGRGLKKLQ-----IPIEK---KDIITHLSQEEK- 349
Db 1591 VIEELLENYDS--SEENNNNLSQKQKLKELTJNFMAEIKKIDKIIKENLDLKLLETNRN 1649
QY 350 -----ELIKRIQSDSLSTE--EKEFLKKLQIDRDSISEEKEL-----LN 391
Db 1650 CMLFTHHTLAEYKIRIDYKFISSATKFSKEFLKYIG-DISNSINDIATLQKLYDHA 1708
QY 392 RIQ-----VDSNPLSEKEKEFLKKLKLIDQIOPYDINORLODPTGGLDPSINLDV 441

Db 1709 QINKVYTSKLSDATNDNNNLIENEKE-----ATQAIKNLRLFTLIDSNUNIDA 1755
QY 442 RKQYKRDIONI--DALLHOSIGS--TLYNKIYXEMNI----- 476
Db 1756 NALHNKKIQMVFNSELHSIESIKOLYKRMHVKLLNIGQINGKYFDISKQFDNLQLO 1815
QY 477 -NNLFWTLGADL--VDSNTNKTINRGI-----PNEFKKFKYSISSNMYIVDNE 523
Db 1816 ESELTANLN-DLKEIGOKISDKKKNKFLHALNENPIFNFTKIEYIHYDYKRRQIDEIEN 1874
QY 524 RPALDNE-----RLKWRIQLSPDTRAGYLENGKLILQNLNLEIKDVOIIOSEK 573
Db 1875 ITSEENENITLYIDITTKLEKQVOSILNFVTEYENDSIQKHIDOPTNENDVSKIESLK 1934
QY 574 EYIRIDAKVYK-----SKIDTIQEQQLNINQEW--NKALQL 609
Db 1935 TTIOQFOELTKINGIKAPYDNNNNINNTISTISQOVNDVKKHTSKDLTENLEIET 1994
QY 610 PRYTKLI--TFNV--HNRYASNIVESAYLLILEMKNNIO--SDLIKVTNVLVDGN 659
Db 1995 QKSLEDIKNSTYEIRGANNNYVATINRYEQQTNRKQNNNSKDEIDDILOKILNTYKKE-- 2052
QY 660 GRVFFDITLPNIAEQYTHODELYEYHSGLYVPESRSLIHGPKGY-----ELRND 713
Db 2053 -----SETRLPITGKNKNVTSIISRI-NKVINLIESE-----YGNNNNYSYNAKKEED 2102
QY 714 SEGFIHEPFAHVDYAGYLLDKNOS--DLVYNSKKEFIDFK-----DEGSNLTSGRT 764
Db 2103 ANSTIID-----LDKSQNLKLDLQOONLKTIIDDLKKNKOEIENRNMLQITNRE 2150
QY 765 NE 766
Db 2151 QE 2152

RESULT 4
Q9BJX9
ID Q9BJX9 PRELIMINARY; PRT: 2747 AA.
AC Q9BJX9;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 235 KDA RHOPTRY PROTEIN (FRAGMENT).
GN IIA.2.
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-YM:
RA Khan S.M., Jarrar W., Peter P.R.;
RT "Distribution and Characterization of the 235 kDa Rhoptry Multigene
RT Family within the Genomes of Virulent and Avirulent Lines of
RT Plasmodium yoelii.";
RL Mol. Biochem. Parasitol. 0:0-0(2001).
DR EMBL: AF323443; AAK15626.1; -
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 2747 AA; 321237 MW; 430508AC7B8824BD CRC64;

Query Match 6.4%; Score 254.5; DB 5; Length 2747;
Best Local Similarity 18.8%; Pred. No. 0.011;
Matches 196; Conservative 168; Mismatches 298; Indels 379; Gaps 44;

QY 45 KEKKNKDNKKRDEERNK-----TOEHLKE---IMKHIVIEVKGEEAVKKEAERK 94
Db 1648 KQKNNEIDQKKEDEVNSKIKNIEMTVNQHKMYEIGIYEKINEIAKTNKNI--ESTRE 1705
QY 95 LLEKVPDVLKMYKAIG-----GKIYVDGDIYH-----ISLEVL 130
Db 1706 LKPTIQHIISSFNAXDLEGIDSDENLGRKXXTEKXGNIYEERIKSYNLITNLETYSKESI 1765

QY 131 SEDKKIKIDYKDALHEHVAKGEPEVLYIOSSDYVENTE-----KALNV 181
DB 1766 TYNOIQMKRIDTOKELL-----KNIEVNNKASLYDIKENEFDRIYTHFKKIKNTV 1817
QY 182 YEIGKILSRDLISKINOPYOKFLDVLNTIRKASDSODDLFTNQKE-----HPTDFS 235
DB 1818 NDNRK-----NEXSVNNGFNDISINSITVKNSTDBNSL-LNLTNQTMEKANIYNNYYIS 1872
QY 236 VEFLEONSNEVOEVEAK-----AFAYIEPQHRDVL--- 266
DB 1873 YKY-----EAENIFRNIPKLANTLNIRKSSGIDLFKDKIKIILSYLDSKEDTLIFI 1926
QY 267 -----QYAPAEAFNY-----MDKFEQELNLSLE----- 290
DB 1927 PSPQKTEYTKISDSYSLIDLIIKKSQEOQKQOFTKLFEFRRLRYEKVQATNELRG 1986
QY 291 -----ELDKQMLSRYE-----KMEKIKQ--HYOHMSDS 317
DB 1987 LSDKYYKKKILSEVKLLIKKSNELKSCNFQNYDTILESCKYDQYKESNNYXQEK 2046
QY 318 L-----SEEGRLIKKLQ-----IPEP 335
DB 2047 LGIDFNADMEKENNDIKVIEELENVYDSEENNIIOSKQIKELTNKFXAEIKRID 2106
QY 336 K-----KODIISLSQEEK-----ELIKRIQIOSSDFLSTP---EKPEFLKKQIDI 378
DB 2107 KXIEKNQDLIKLETRKNCMLFHTTLAETLKIKITYSKIESATVFSSEFLAYIG-DT 2165
QY 379 RDSLSSEKEL-----LNRIQ-----VDSSNPLSEKEKEFLKLLKIDQYDINO 423
DB 2166 SNSLNDIATQKLYDHLQNKVYTSKLSDATNDNNLIEKE-----ATQ 2212
QY 424 RLDDTGLIDSPSINDVRYOKRDIONT--DALLHOSIGS--TLVKKIYLYENMTI--- 476
DB 2213 AIRKTLFTLFTDSNNIDANALHNKNIQMYVFNSELKHSIESIKOLYKMHVFKLLNTGOI 2272
QY 477 -----NNLTATLGADL--VDSDTNFKINRGI-----FMEFK 505
DB 2273 NGKFEDISKOPDNILQOSELELTANLN-DLKEIGQISDKKFKLHMLNETPIPNFTLK 2331
QY 506 KNFYSISSNMYVIDINERPALDNE-----RLKWRITQSPDRAGYLENGKILIQ 555
DB 2332 EITHDIYKRYKQIDELINISSEMENTIYIDITTKLKEVQOSILNVVTYENDSNIIKQ 2391
QY 556 RNIGLEIKDVOITKQSEKEYIRIDAKVPR-----SKIDTKIQE 594
DB 2392 HIODTNEVDYSKIKESLTKTIOSEIINKINGIKAOFYNNNNINININTISTISQVND 2451
QY 595 AOLNINQEW---NKALGLPYTKLI---TFNVNHRVASNIYESAVYLLNEKKNIQS--- 645
DB 2452 VKKHISDOLTXEBELIETQSLIEDIKNSTYEIKSEQITNTVNTIRNVEEQOTNKIUNNSN 2511
QY 646 -----DLIKRVNTLVGNGRFVFTDITLPNIAEOYTHODEIYBOVHSGLYVESRSIL 700
DB 2512 KDEIDITIQILVNNKE-----SETKRLPTTGKNNVYISIRI-NKYVINLSE--- 2560
QY 701 LHGPKSV-----ELANDSEGFHEFGHAVDYAGYLLDKNOS---DLVTSKKFLDIF 751
DB 2561 -YGNNNNVSYNAKKLEDEDANSIIDL-----LKSQNILDLIOONIKIIDDL 2607
QY 752 K-----EEGSLNLTSGRTNE 766
DB 2608 KNKKOELERNNLQITNREOE 2628
RESULT 5
096133 PRELIMINARY; PRT; 1979 AA.
AC 096133;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, last annotation update)
CP HYPOHETICAL 237.7 KDA PROTEIN.

GN PRB0145C.
OS Plasmidium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tetteh H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shalton S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lal Z., Schwartz D.C., Perlman M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "chromosome 2 sequence of the human malaria parasite Plasmodium
falciparum";
RL Science 282:1126-1132(1998).
DR EMBL; AE001375; AAC71819.1; -
KM Hypothetical protein.
SQ SEQUENCE 1979 AA; 237745 MW; 5C6CDB307AF7D37 CRC64;
Query Match 6.3%; Score 253; DB 5; Length 1979;
Best Local Similarity 19.6%; Pred. No. 0.0083;
Matches 158; Conservative 163; Mismatches 253; Indels 234; Gaps 38;
QY 42 MHVKEKKNNDKRDDEKRNKTOEBHLKEIMKIIVIEYKGEAVYKKEAEKLEKVP 101
DB 329 LNKQEKEREKEREKEREKEREKEDYDLIK-----ELKDEKI-----SILEKVS 376
QY 102 DYLEMYKAIGKIYIVGDITK--HISLEALSEDKKKIKIDYGDALHEHVYAKGEYE 159
DB 377 -----IVREMDIEKRHNFLH-MEDDLKOLKNSFYANN--NOLAKYKCE--- 418
QY 160 PVLVIOSSDYVENTERKALNAYVEIGKILSRDLISKINOPYOKFLDVLNTIRKASDSGQ 219
DB 419 -----IKMLTELEKEKELK---DIENVSKKEIKLNQNEKEKQIIFENKKNHE- 468
QY 220 DLFTNOLKEHPDPSVPELEONSNEVOEVPKAFAYIEPQHRDVLQYAPAEFNM 279
DB 469 -----IHGLKEELKE-SVKIRKIEFQELQEM-----VDIKQKEDQL- 507
QY 280 FNEQINLSLEELKDOQMLSR-----EKMEKIKOHYOHMSDSLSE- 321
DB 508 YMOIESISIELSKKEKEKYNOYKNTYIEBINNLNEKLEETKKEKETTINQNTTINMLNN 567
QY 322 ---GRGLKKLQPIPEPKDDI-----IHSLSQEEKEL---LKRQIDSSDFLSTE 368
DB 568 DIHMLNNGIKTMNQISTLKNDVHLNEQIDKLNNKGTLSKISELVQIMDL--KEEK 625
QY 369 EFLKKLQIDRDSLEBEKELNRIQVDSNPLSEKEKEFLKIKL----- 414
DB 626 DFLNQIIVDSNQI---DLTKRMEKEKMLEQENKRYKQEMELLRGNKSSENILNN 680
QY 415 DIQPYDINORL-----ODTGLIDSPSINDVRYOKRDIQNDIALHOS 459
DB 681 DEVCYDLKRLKSKESKMKMKKEHDKKLAELKDCOVR--IREMKNKNEKIMLKEE- 737
QY 460 TGSTLYKVIYLYENMTNNLTATLGADLVSDNTK--INRGIFNEFKNFRYSISS--- 514
DB 738 -----YED-KIN---TLKQONEDKINTLKQONEDKINTLKVEYEHKINTKKEE 781
QY 515 -NYMIYVINERPAID---NRLKWRITQSPDTRAGL-----ENCK 551
DB 782 YEHKINTLNEONENKINTLNEONENKINTKVEYEDKMNLTNEONEDKMSLKEEYENKI 841
QY 552 LILORNIGLEIKDY--OLIKOSEKEYIRIDAKVPRKSIDPKIOAOLNINO----- 601
DB 842 NQINSNNEIKIDVYNEYIEEVDKLKATLDEK---KKQFKELIYAIKRAHEKQILLTE 898
QY 602 -EWNKALGPRTYTKLITFNVNNRASNIVESAVYLLN-EKMKNIQS--DLIKRVNTLVLDG 658
DB 899 MEELKQCRDKKYSGL-----YEKTI-KLISICQITINIECCDDIENDDIIRIEEYINN 952
QY 659 NGRFVFTDITLPNIAEOYTHODEIYBOVHSGLYVPESRSILLHGPKGYELANDSEGFI 718

```
Db      953 KLL-----KKEVEEKKER-          HSSFNLTKSEKR-   979
Oy      719 HEFGHADVADYGLDKNOSDLVTSRK    746
        |::|::|::|::|::|::|::|::|::|::|
Db      980 --FKNSIEDSKSHELKKRKHEDLSKDKE  1005

RESULT           6
Q26216            PRELIMINARY; PRT: 2771 AA.
AC Q26216;
CD Q26216;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DE 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DEF RHOPTRY PROTEIN.
OS Plasmodium yoelii.
OC Eukaryota, Alveolata, Apicomplexa, Haemosporida, Plasmodium.
NCBI TaxID=5861;
RN [1]
RP SEQUENCE OF 379-2771 FROM N.A.
RC STRAIN-YM;
RX MEDLIN-97077455; PubMed-8920022;
RA Sinha K.A., Keen J.K., Ogun S.A., Holder A.A.;
RT "Comparison of two members of a multigene family coding for high-
RM molecular mass rhoptry proteins of Plasmodium yoelii.";
RL Mol. Biochem. Parasitol. 76:329-332(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-YM;
RA Green J.L., Holder A.A.;
RT "Structure of the EB gene encoding a high molecular mass rhoptry
RM protein of Plasmodium yoelii.";
RL Mol. Biochem. Parasitol. 0:0-0(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-YM;
RA Holder A.A.;
RT Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U36927; AB41263.3; -.
SO SEQUENCE 2771 AA; 325640 MW; CCCCBA6BE7AFCE36 CRC64;

Query Match             6.3%; Score 253; DB 5; Length 2771;
Best Local Similarity   20.2%; Pred. No. 0.013;
Matches 182; Conservative 159; Mismatches 340; Indels 222; Gaps 38.
```

```

Db      2093 KHESDNNFSEBENNIIQSKKKKMLKELTFAFAEIKIEDKIIEKNGLINKLIERKDDM    2152
OY      351 -----LKRIODSSDF-----LSTEKEFLAKQIDIRSLSEPEEKEL-----LNRI    393
               ||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      2153 LFTYKTLVEFLKLTIDYKRTITSATKSFKSELKIDA-TSNSLDNDINTLOTKYDLNOI    2211
OY      394 Q-----VDSNPLSEKEFEKLKLKLDLPDYDINQRLODPGGLDSPIMLVPRK    443
               :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      2212 NKHVASWADARINDNNNNLLERKEKTKTINNLTELF-----TIDSKKIADGGH    2260
OY      444 QYKRDIQNIDLALHOSIGS--TLNYKIYLENMNT-----NNTATLGDVL    487
               ::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      2261 NNKIDIIYFNSELHKISIDSIKOLYKRMHAFKLNLIGHINKKFYDISKEFDNILOLES    2320
OY      488 VDSNTNKIRNGIRFENPFKKNFYSNSMYIDIEMERALDNERKLRTOLESPDRAGYL    547
Db      2321 TANLMDLEIGOKISDKRQPHALSEP-PIRENTELTKEITHYDIYKKQIDIEINITNE    2379
OY      548 ENGLILORNICLET-----KDVOITIKOSEKEYIRIDAKVPKSRIDTRI    592
               ||::||::||::||::||::||::||::||::||::||::||::||::||
Db      2380 EMENTTLVMDITTKLMKKKVQSILNPVTTYENDSNIIKHQIDONNENDSKI-KESLEPTI    2438
OY      593 QEAQLINQENKKAALGPLRYTKLTIFNVNHNRASINAVESAVALILEMKNNIOSDLIKVYT    652
Db      2439 QSFQKILN-----KLMEIKAQFYDNNNNINNAVISTISODVDNPKHSIKDE--TIE    2486
OY      653 NYLVQNGRFVFTDTLPLRIAEOY--THODEYEOWHSGVLGVPSRSILLGPCKGYEL    710
Db      2487 NELIQ-----IQKSLEDIKSTYDINSEQITTKTVNPLHDYVEOOTKIQNNPMK----    2535
OY      711 RNDSEGFIHFEGHAADVAYGLLDKN-QSDL---VTNSK-----KTFIDFREE    754
Db      2536 -----DEIDLIOEIYNVNKESSELKLPILNNKDVPTLISRDKVINLIKSE    2583
OY      755 GSN 757
Db      2584 YNN 2586

RESULT       7
O9BK46 ID O9BK46 PRELIMINARY: PRT: 3130 AA.
AC O9BK46.
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE- RETICULOCYTE BINDING PROTEIN 2 HOMOLOG A.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_taxid=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21101060; PubMed=11160005;
RA Triglia T., Thompson J., Caruna S.R., Delorenzi M., Speed T.,
   Cowman A.F.;
RT "Identification of proteins from Plasmodium falciparum That Are
   Homologous to Reticulocyte Binding Proteins in Plasmodium vivax.";
RL Infect. Immun. 69:1084-1092(2001).
DR EMBL; AF312916; AAK19244.1; -
SQ SEQUENCE 3130 AA; 370415 MW; 13D973DB89BD2026 CRC64;

Query Match          6.2%; Score 248.5; DB 5; Length 3130;
Best Local Similarity 19.8%; Pred. No. 0.022;
Matches 184; Conservative 146; Mismatches 274; Indels 325; Gaps 43;
```

QY	129	ALSDKKKIKDI	-----YGR-----DALHBNY-----YUAREGEVPLVIOSSSDYENK	176
Db	1803	LDYMEIKKEIVQAOTEDKLKAYDEL	-KNYLNINKEBGNHLL-----DYIK--EK	1851
QY	177	ALNAYVEIGKILSRDLS	-----KINQYQK-----FLDVLNTKINAKSDSDGDLF	223
Db	1852	IFNLYIKCS	--QONITDSDYNTYVAKQYKTIETDVKFLDLSLNTIEKKNSVANLEIC	1909
QY	224	TNQ	-----LKEH-----PIDSVEFLQNSNEVOEYFAKAFYTYEPQH----	262
Db	1910	TNKEEDIKNLKLVIKILANFS	GIIVMSDPTNETPREPDLNPLL--NLQYPERKHETTS	1967
QY	263	-----	RDVLQ--YAPAFVNYMKFN-----	282
Db	1968	TLENDSDLELDHGSNDESDIDMLKYYNDILELHTYSTOILKTYLNDIOKLKDCDCLYKD	2027	
QY	283	---OEINLSLEELKQDRMLSYREKMEKIKOHYQHSWDSLSEBGRLLKLLQIDPIEPK-	337	
Db	2028	CKELRELSTALYDKIQ--TTSVINRENDISNNIDYSNKLTNE-----IDAIOYNEKYE	2081	
QY	338	-----DIIHLSQSEKELLKRIQIDSS-----DFLSTEKEFLKQIDIRDSL	382	
Db	2082	IFDNVEEYKTLDDTKNAYVYKKAETIKLNDINKTKEDDLIYENDIDELKSL-----TL	2135	
QY	383	SEEEKELNRIQVDSNPLSEKEKEFLKLKLDIQDYNORLODPTGDISPSINLQVR	442	
Db	2136	SSNMEIKTIYQ--NSYSPS-----DLINKIN-----DID	2164	
QY	443	KQYRDIQNDIALHQ--SIGSTFLYKTYLYENMNINLTATLGADL--VDSTDN----	493	
Db	2165	KEMKTLIPMLDELLECHGNIDISLYNFI-----IRNIQIKGNDIKINIREQENDTIC	2217	
QY	494	-----TKINRGTFENFKKNFYKS--ISSNVIYV-----INERPALQNR	531	
Db	2218	PEYIQNNYNFLKSDISLFNKYDDHIKVDYTSNNIDYVKNHSLSEHYATNATITEN-I	2276	
QY	532	LKMRIOISPTRAGYLENGK--LLIQRNIGLEIKQDVQIIOSEKEYIRIDAKVYPKSKI	588	
Db	2277	MTSIVEINEDPMNSLEETQDKLLELYENF-----KKEKNIINNKYIVHFKL	2322	
QY	589	DTKIQ--EAOINQNMKNALGLPKTYKLITTNVNNRASNIVESAYLILEMKNKNIQS	645	
Db	2326	-KEIENSLEYTNSITFNFKI-----METQNDILKNEF--NNKT	2363	
QY	646	DLIKYVNYLYVGDGRVFPDITL-----PNIADQYTHODEIYQVSK	689	
Db	2364	KINDKVR-----ELVHYDSTLTLESIOFTNNLKYGLDMSNIQDYKEDIENNVELKV	2415	
QY	690	GLVYEPSRSILLHGKSGVGLRLRDSSEGFIEHFCAVDYAG-----YL-LDKMSQDVL	741	
Db	2416	KLYIENTNLL-----GRINTFKELDKYQDENNGIKYIEINKENSYI	2466	
QY	742	TNSKKFLIDIREGSGNLTSGRTNEAEPF	770	
Db	2461	IKLKEKANNLKENFSKLLQNIKKNETELY	2489	
RESULT 8				
Q9BK45	PRELIMINARY;	PRT;	3254	AA.
AC	Q9BK45;			
DT	01-JUN-2001 (Tremblrel. 17, Created)			
DT	01-JUN-2001 (Tremblrel. 17, last sequence update)			
DT	01-JUN-2001 (Tremblrel. 17, last annotation update)			
DE	REICHLUCYTE BINDING PROTEIN 2 HOMOLOG B.			
OS	Plasmodium falciparum.			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=5833;			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21101060; PubMed=11160005;			
RA	Triglia T., Thompson J., Caruana S.R., Delorenzi M., Speed T.,			

RA Comman A.F.:
RT "Identification of Proteins from Plasmodium falciparum That Are
RT Homologous to Reticulocyte Binding Proteins in Plasmodium vivax.",
DL Infect. Immun. 69:1084-1092(2001).
DR EMBL: AF312917; AAK19245.1; --
SQ SEQUENCE 3254 AA; 382876 MW; 6F9CAFA2AA6167BA CRC64;

Query Match	6.2%;	Score 248.5;	DB 5;	Length 3254;
Best Local Similarity	19.8%;	Pred. No. 0.023;		
Matches 184;	Conservative 146;	Mismatches 274;	Indels 325;	Gaps 43;

QY	44	VKEKREKKDKDRERNTQEHLEKMHVKEVGEAVKAEAK	-----	95
Db	1684	LKEQNNAEYKKMNEYVNVNTE-KEIHKHKNYEIRIMHKEKNEKKKMSNS	1742	95
QY	96	-----	LEKVPDLEMYKAIGKIYVDGDTKHISLE	128
Db	1743	KSLLTLDSPSRMEYNEYINDYINENFENHQHILNIEJNGFESYNIINIKMEIIND	1802	1
QY	129	ALSDKKKKINDI-----YK-----	DALLHEHY---YAKEGEPELVLISSSDYENNEK	176
Db	1803	LDVYEKEIEVQOTEDYDLNKKVDEL-KNVLNNIKEDGRLI-----	DYIK-EK	1851
QY	177	ALANYEIGKILSRDILS-----	KINOPYOK-----FLDVLNTKNASDSGODLLEF	223
Db	1852	IFNLYIKCS-----QOONIIDDSYNTVAKOYIKTIEDYKFLDLSMTIEEKNKSVANLEIC	1909	
QY	224	TNO-----LKEH-----	PYDSEVELEONSNOVOBFKAFAYAYEPOH-----	262
Db	1910	TNKEDIKNLKHVYIKLANSGIIIVMSDITNIEITPEPPLDNDLL-----	NIQLYFERKHETTS	1967
QY	263	-----	RDVLOL-YAPAFMYDKFNE-----	282
Db	1968	TLENDSDLELDHGSNDESIDNLKAYNIDIEHTYSTOILYLNIOIKLGDCNDLYKD	2027	
QY	283	-----OELNLEBELKOORMLSREKKEKIKONQHMSDLSSEGGGLKKTIOPIERPK-	337	
Db	2028	CKELRELTALYDKTIO-TTSYVNRNDISNNIDIVSNKLN-----	IDAIQYNEKYKE	2081
QY	338	-----DDIHSLSOEKELLRIOIDS-----	DFLSTEKEFLKLQIDIRSL	382
Db	2082	IFDVEYEKTLTDPTKNAYIKVAKIEILKNVDINKTKEDLDIYFNLDLEKSL-----	TL	2135
QY	383	SEEKELLNRIOVDSNPUSPEKEKEFLKLKLDIOPDIQNRLOOTGGLSDPSJNLNDR	442	
Db	2136	SSNMEIKETIVO-NSYNSFS-----	DINKNIN-----	2164
QY	443	KOYRDIQNDILHOO-SIGSTLYNKIYLEMNMINNTATLGADL-VYSTON-	493	
Db	2165	KEMKTLIPMLDELBESHNIDISLYNFI-----	IRNIQIKGIDKNIREOENDTJIC	2217
QY	494	-----TKINRGIFNEKKNFYAS-SSSNVYIVD-----	INERPALDNR	531
Db	2218	FEYIONNYNFKSDSIFNKYDHIKVDYVNSINIDVYKHSLSLSEHYNTATNIEN-I	2276	
QY	532	LKMKIOLSPTRAGYLENGK-LIIQNRNGLEIKVQIIRKSEKEYIRIDAKVYPSKI	588	
Db	2277	MTSIVEINEEDENUSLEETODKLLEYENF-----	KKENNIINNKKIYHFNKL	2325
QY	589	DTKIO-----EAOQNIINOEWNKALGPYTKLITFENVHNRASVIVESAYLILNEMKNJOS	645	
Db	2326	-KEIENSLELEYNSJSTFNKI-----	NETONIDILKNEF-NNKT	2363
QY	646	DLIKKVTNYLVGNGRFVFDITL-----	PIIAEOYIHODEIYQVSHK	689
Db	2364	KINDKRV-----ELVHVSDFLTLESIoTFNNILYGLDMSNIODYKVEDINNVELKV	2415	
QY	690	GLYVPEBSRILHOPSKGVELRNDSDGFIHEPCHAVDAG-----	YL-LDKNQSOVL	741
Db	2416	KLYIENTITNLL-----	GRINTFKELDKIQDENNGIKDIEIENKENSNT	2460
QY	742	TNSKKFIDIKREGSNLTSYGRTEAEFF	770	

DB 2461 IKLKEKANNLKENFSKLLQNIKRNTEELY 2489

RESULT 9

09G276 PRELIMINARY: PRT: 1387 AA.

ID 09G276 AC 09G276:

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE RETICULOCYTE-BINDING PROTEIN 2 HOMOLOG A (FRAGMENT).

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.

NCBI_TaxID=5833;

RM [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20402589; PubMed=10920203;

RA Rayner J.C., Galinski M.R., Ingravallo P., Barnwell J.W.:

"Two Plasmodium falciparum genes express merozoite proteins that are related to Plasmodium vivax and Plasmodium yoelii adhesive proteins involved in host cell selection and invasion."

RL Proc. Natl. Acad. Sci. U.S.A. 97:9648-9653(2000).

DR EMBL; AF196347; AAF98066.1; -

FT NON_TER 1

FT SEQUENCE 1387 AA; 164871 MW; 5660154744F0C285 CRC64;

Query Match 6.2%; Score 245.5; DB 5; Length 1387;

Best Local Similarity 19.9%; Pred. No. 0.011;

Matches 186; Conservative 143; Mismatches 269; Indels 337; Gaps 43;

DB 44 VKEKKNDENKRRKDEERNKQOEHLKIMKIVKIEVKEBAVKKEAEKL----- 95

DB 48 LKEKONMEAEYKKNEMNYVETE-KEIKKKKVEIRIMEHIKKEKKKKKMESSNN 106

DB 96 -----LEKPSDYLEMKKAIKGGKIYIDGDKTKISLE 128

DB 107 KSLTLMDSFRSMFYNEYIINDYNINENEFKHOITLNEITNGNESYNIINTMTETIINDN 166

DB 129 ALSEDDKKIKDI-----YGR-----DALLHEHY-----VYAKEGEYPLVIOSSDYVENTEK 176

DB 167 LDYNIKEIKKEVAQTEYDKLNKKVDEL--KNYLNKIKEGEGRLL-----DYIK--EK 215

DB 177 ALNYYEIGKILSRDILS-----KINOPYQK-----FLDYLVNTIKASNS----- 216

DB 216 IFNLTKGSE--QONIIDDSYNYIVTKOYKTIEDVKFLDLSLNTIEKKNSVANLEIC 273

DB 217 -----DGOPLL-----FTNQL 227

DB 274 TSKEIKNLKLVIVILANFSGLIWSDTNTEITPENPLEDNDLNLQILFEKKHETISFL 333

DB 228 KEHPDPSVEFLQNSNEVEYFAKAFAYIIEPOHVDLYLYAPEAFNYMDKFN----- 282

DB 334 -ENDSDLELDHGSNSDESIDLN-KVYNDILE-----LHYSTQILKLDINQIKGDC 385

DB 263 -----OEINLSLEELKQDRLMSREKWEKIKOHYQHSDSISEGKGLKQIP 332

DB 386 NDLYVCKELRELSTALYDLKIO-ITSVINRENDISNNIDIVSNKLN-----IDAIOYN 439

DB 333 IEPKK-----DDIHSLSQEEKELKRIQIDSS-----DELSFEKEFLKKLOI 376

DB 440 FEKYEIDDNVEEYKTLDDTKNAYIVKKAELLKNDINKTKEDLDLYENDLDELEKSL- 497

DB 377 DIRDSLSEEEKELLRIOYDSSNPLSEKEKEFLKLLDIOFYDINQRIODTGGILDS 436

DB 498 -----TSSNEMKIKTIYQ-NSYNSFS-----DINKNN- 525

DB 437 INLDVARKOYKRIQONIDALLHO--SIGSTLYNKKIYLYENMMNINNTLATGADL--VDSFD 492

DB 536 ---DIDKEKMTLIPMLDELLNEGHNIDISLVNFI-----IRNIOIKIGNDIKNIROE 575

DB 493 N-----TKINRGINEKKKNFKYS--ISSNYMIVD-----INERP 525

DB 576 NDTNICEFYIÖNNYNEFIKSDISIFNKYDHIKVDNYIISNNIDVANKNSLSSEHYIATFN 635

DB 526 ALDNERLKWRIOLOSPDTRAGYLENGK---LIÖRNIGLEIKDQVILKÖSEKEYIRIDAKY 582

DB 636 ILEN-IMTSIVELNETENSLEETQDKLELYEN-----KKEKNIINNMYKI 683

DB 583 VKRSKIDTKIQ---EAÖLINOEWKALGLPKYTKLITFNVHNRRASNIVESAYLILNEM 639

DB 684 VHFNNK-KEIENSLEFYNSISTNFNKI-----NETONIDILKNEF 722

DB 640 KNNIOSDLKKYTNVYDNGRFVFTDIL-----PNTAEQYTHODELY 683

DB 723 -NNIKTKINDKVK-----ELVHVDSTLTLESIOTFNNLGYDLMSNIÖDYAKIEDINN 773

DB 684 EQVHSGKLYVPESRSILLHGPSKGYELRNDSEGFIEHFCNAVDDYAG-----YL-LDK 735

DB 774 VELKKVKKLYIENTNL-----GRINTFIKELDKYGDENNGIDKYEINK 818

DB 736 NOSDLVTNKKKFLIDTFKEGSLNLTSGRTNEAEFF 770

DB 819 ENNSYIIRIKLEKANNLKENFSKLLQNIKRNTEELY 853

RESULT 10

09YVT6 PRELIMINARY: PRT: 1127 AA.

ID 09YVT6 AC 09YVT6:

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE ORF MSV156 HYPOTHETICAL PROTEIN.

GN MSV156.

OS Melanoplus sanguinipes entomopoxvirus (MsePV).

OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;

OC Entomopoxvirus B.

OX NCBI_TaxID=83191;

RM [1]

RP SEQUENCE FROM N.A.

RC STRAIN=TUCSON;

RX MEDLINE=99102612; PubMed=9847359;

RA Afonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.:

"The genome of Melanoplus sanguinipes entomopoxvirus,"

RL J. Virol. 73:533-552(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=TUCSON;

RA Afonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.:

Submitted (May-1998) to the EMBL/Genbank/DBJ databases.

DR EMBL; AF063866; AAC97677.1; -

FT SEQUENCE 1127 AA; 134265 MW; F185DAID5A3FE7D1 CRC64;

Query Match 6.0%; Score 239; DB 12; Length 1127;

Best Local Similarity 20.8%; Pred. No. 0.016;

Matches 159; Conservative 121; Mismatches 252; Indels 234; Gaps 35;

DB 60 ERNKQOE---HLKEIMKIVKIEVGEBAVKEAEKLEKVPDLEMYNAIGKIV- 115

DB 24 ENNKVSLDIINSLEYELNNI-----KFSDKRTNIIKINNYKIVKIEKIF 66

DB 116 -----IYGDITKHSLSALSEDDKKIID-----IYGDALLHEHYVAKSEYEV 161

DB 67 MHNQFKINDYNIILQYLIYEYNEINCKIKENFPCKNPLY-----NITYKKKLYIAD 118

DB 162 LVYQSSDYVENTEKALNYYE-----IGKILSRDILSKINOPY-----QKLDVYL 207

DB 119 L-----DYEEKDKELVYNIEOKNAVDKI--NDIKNNVNHSHDETITTKETLIDIL 170

DB 208 NTKANSDSDGODLFTNOLKEHPDPSVEFLQNSNEVEYFAKAFAYIIEPOHVDILO 267

DB 171 NKLLKLVSSDEKQOLI--EQIYKNNKKEIEF--KNIDNVQKEINK----- 211

```
Qy 268 LYAEAFNVMKREBOEINLSLEFLKQDRMLSRKEKWEKIKOHQMSDSLSEGRGLK 327
      : : : : :
Db 212 -----KODELNKLDE-----SKKEFK----- 229
Qy 328 KLOIPFKKDDIIHSOEEKELKRIOIDSSDFLSTEEKEFLKQIDIRSLSEEEK 387
      : : : : :
Db 230 -----KODELNKTIDKQOEFLIKLNDKEINFINIDEKOKLDDQINSKI-ITLLENIK 280
      : : : : :
Qy 388 ELLNRIOVDSSNPLSEKEKEFLKKLIDIOPIYDINORLODTGGLIDSPSINLDRKOKYR 447
      : : : : :
Db 281 GVMN-LYTERKKNISNLOEINLKN-DSTIKSLDEKOKLDD-----ELDK 322
      : : : : :
Qy 448 DIONIDALLHOSIGSTLYKNTIYLENNMNINLRATLGADLVDSIDTNTKINRGIFNEKKN 507
      : : : : :
Db 323 NINNI-----TSLYKN-----SNRKITNIOQLLESSLTD-FNNANIN-----INELSK 365
      : : : : :
Qy 508 FKYSISSNYMVIDNERPALDNERLKWRIOLSPDTRAGYVLENGKLIQRNIGLEIKVOI 567
      : : : : :
Db 366 IK-----LFD-NDQKLANDITEONKKT-----DFNNSTRIRKEKLDTEYKKIDD 411
      : : : : :
Qy 568 IKOSEKEYIRIDAKVYPKSKIDTRIOQAOLINQEWNKALGLPKYKLTIFENVHR----- 623
      : : : : :
Db 412 IKNNNNLOKLEESYK-----KIDEOTEYKKNKINKEYNDIEL-KNNMLQKLEENKKIDE 465
      : : : : :
Qy 624 ---YASNIVESAVLILNEMK-NNIQ-----SDLIKVTNYLVDCNGRFVFTDITLP 670
      : : : : :
Db 466 QTEYKKNKIKETNDITELKNNNNLOKLEENKKNINDKLTILKNDIESNTLF-----NKL 520
      : : : : :
Qy 671 NIAEOYTHODEI-----YEOVHSKGLYVPESSSILHGFSGVLENDSEGFIERGHA 724
      : : : : :
Db 521 NIDFKOKSREIKLKNTEYQOL-----KRDLEENKNTLMLKSLDNKLSLEQL 570
      : : : : :
Qy 725 VDDYAGYL--LDKNQSDVLVNSKKFLDIFKEEGSNLTSYGRINEAE 768
      : : : : :
Db 571 YDSKKNILDCIDKIYNSLKEKNKIDEXF---SNIEKFIYVYIE 612
      : : : : :

RESULT 11
Q9EMP3 PRELIMINARY; PRT; 1238 AA.
AC Q9EMP3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE AMV156.
GN Amsacta moorei entomopoxvirus (AmePV).
OC Viruses: dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
OX NCBI_TaxID=28321;
RN SEQUENCE FROM N.A.
RX MEDLINE=20396580; PubMed=10936094;
RA Bawden A.L., Glassberg K.J., Digans J., Shaw R., Farmerie W.,
RA Moyer R.W.;
RT *Complete Genomic Sequence of the Amsacta moorei Entomopoxvirus:
RT Analysis and Comparison With Other Poxviruses.*;
RL Virology 274:120-139(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Bawden A.L., Glassberg K.J., Digans J., Shaw R., Farmerie W.,
RA Moyer R.W.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF250284; AAG02862.1; -
SQ SEQUENCE 1238 AA; 147096 MW; 9BBF80A39DB6E35B CRC64;
```

Query Match 5.9%; Score 237; DB 12; Length 1238;
Best Local Similarity 20.1%; Pred. No. 0.022;
Matches 189; Conservative 170; Mismatches 293; Indels 290; Gaps 52;

```
Db 158 RSODEFINIKNTYHDIKN-----DNNIIDQLSRPIDITILKNEFLDKLT 205
      : : : : :
Qy 115 -----YVDDGITRKHIS-----LEALSDDKKIKIDYKQALLHEHYV-----AKEG 157
      : : : : :
Db 206 NYIKNTYISDDKLVNIISKNPLFLNNIKIKEYINENKINNIILNK-IYVMDDDLFIYNNVKN 264
      : : : : :
Qy 158 YEPVLVQSSSEDEVENTKALNYYEIGTLSRD-----ILSK-----INQYOKF 203
      : : : : :
Db 265 YDEINLEFKNNEDEFLIFLKNITINDY-INNILLSSEYKNIIDMEIKSKNIDIIMDYVEKN 323
      : : : : :
Qy 204 LDVINTIKNSDDSG--QDILFTNQ--LKEHPIDFSVEFLEONS-----NEVQ 247
      : : : : :
Db 324 SDVATKLLQKISDEYLSNILLSNENFLKIKOIISSENIDISLNSLNIENKFKILK 383
      : : : : :
Qy 248 EVPAKFAVYIEPOHRDVLQ-----YAEAFNVMKRFN--EOEINLSLEELK--- 293
      : : : : :
Db 384 EILNKEFL-----KGNIDMLVNNFNNDYKQYKKSIDEKIDIDINIKODEINTHLSLDVLI 438
      : : : : :
Qy 294 -----DORMLSRYEKWEKIKOHQMSDSLSEGRGLK 329
      : : : : :
Db 439 SPYONLRTYANNIDITMESIYDRYDKOILNLVQETEKLDHVKQNNVSRFRQ-----LSDV 494
      : : : : :
Qy 330 QIPIEPKDD-----IHSLSOE-----KELKRIQIDSSDFLSTEEKEFLK 375
      : : : : :
Db 495 SNSIDEKYDNFKQSAIILKINDKESFYNYVEDISKIKLINDLL-IFKNIIENLN 553
      : : : : :
Qy 376 -----IDIRD-SLSEEEKELLNRIOVDSSNPLSEKEKEFLK-----KLKDIOPYD 420
      : : : : :
Db 554 SGDVNTNIDIKNIDLSEDIKIFINKI-----LPELIKFNLENGEYVNSDKFNIINESV 606
      : : : : :
Qy 421 INQRLDGTGLIDSPSINLD-VKQYKRDQNDALLHOSIGSTLYN-----KIYLENMN 475
      : : : : :
Db 607 IN-----NINTDYIILNITIKSNIPTESSIT-TRVESRLKVIYDKFY-----D 647
      : : : : :
Qy 476 INNLATLGADLVDSPTNTKINRGIFNEFKKNFYSSSYVM--IYDINERPALDNERL 532
      : : : : :
Db 648 INNSISLINTDI--STNNAKISE-IAADIER-IKNNISENIKTIPIEDFDEKAKIDKIL 703
      : : : : :
Qy 533 KWRIOISP-----DTRAG-YLEN-----GKLIIDLRNIGLE----- 561
      : : : : :
Db 704 KY-LEINPVLQSKVYANIKIKEYISYTSRASLOIKNIITILDKILSDNKNINIELSNFS 762
      : : : : :
Qy 562 -----IKDVOITIKOSEKEYIR--IDAKVYPKSKIDTR--IOQAOLINQEWNKAL-GL 609
      : : : : :
Db 763 LDDKIILKSIDLKLNDQDEIKELLSKRTETISSEFKEINAINELLLPLESKNENIAGI 822
      : : : : :
Qy 610 PKYTKLITFNVHRVYASNIVESAVLILNEMKNNIOSDL-----IKVTNYVLVDGNGRF 662
      : : : : :
Db 823 YDEINILIS-KNYNEILFKIDNNMYLOLEKIEENVLDNINVDITIKNNLKEIND-F 880
      : : : : :
Qy 663 VFTDITLPNIAEOYTHODEI--YEOVHSKGLY-----VPESSSILHGP 704
      : : : : :
Db 881 IYKNYSQGNIPKYITETENLRGAETFLNKIIYIYNSLENAPRINFESVPKT-NVSKSK 939
      : : : : :
Qy 705 SKGVELN-----DSEGFIEFGHAV-----DDYAGLILDK----- 735
      : : : : :
Db 940 STYNDVANLIKYNMKEKKNPRLQAGEVOEYKPNILNVIYDPLNKYIKINNTENIKI 999
      : : : : :
Qy 736 -NOS-----DLVNSKKFLDIFKEEGSNLTSYGRINEAEFF 770
      : : : : :
Db 1000 LNNSYKKILFEYDNNNSINDIITK---NIITY-KTEDIISY 1036
      : : : : :

RESULT 12
Q26023 PRELIMINARY; PRT; 1048 AA.
AC Q26023;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOHETICAL PROTEIN (FRAGMENT).
OS Plasmodium falciparum (isolate 3D7).
```

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RA Alano P., Elliott J.F.;
RL Submitted (Apr-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; M69147; AAA74653.1; -
KM Hypothetical protein.
FT NON_TER 1
FT NON_TER 1048
SQ SEQUENCE 1048 AA; 126518 MW; 7AF051480FA22424 CRC64;

Query Match 5.9%; Score 235; DB 5; Length 1048;
Best Local Similarity 19.8%; Pred. No. 0.022;
Matches 174; Conservative 149; Mismatches 301; Indels 254; Gaps 43;

QY 27 FIPVQAGAGHD-----VGMHVKEKKNKDKENKRDENKTOBEHLKEIMKHIVKIE 80
DB 295 FTTISGEENHONRELKKTEANLKEEMKKRFNEQOEQRKKKAE--DENMETIOKHD 352
QY 81 VKGEAVKKEAAKLEKVPDLEMYKAIGKITYVDG---DIRKHISLEALSD--K 134
DB 353 METSKLEKKEEDEVTODEEFD-----KQYAIIDQOELEFLRTDSEGESDVPK 402
QY 135 KRIK-----DIYKDALHEHYVYAKEGEYPLVYIOSSEDYV-ENTEKALNYEYI- 184
DB 403 DKVKPPRPGSPSPFYNNATISSFHEKMELYN--TSISSLNYVKEINKRPDDVYKEIK 460
QY 185 -----GKILSRDIISKINO-----PYQ-----KFLDVLNTIKNASDSG 218
DB 461 SKTYPRFDLTSQTKNCNKLFGKLNETIKDKKEYOKNIOSYKKNVIDIIDIO--KKANG 518
QY 219 QDLFTNOLKEHPTDSVEFLDONSNEVOEVPKAF----- 254
DB 519 KYIIIONLIEIKIDYKGVNRLSDRKFYKNFKYLGKRRKMKLEDFRAQFGAIRFIKD 578
QY 255 -AYTIEPOHROVLO-LYAPFAFNMDKFNEDQINLSLELKDQRMISR--YEKMEKIK 308
DB 579 LTTTWEETTYVLEDIYMEKKKYKEYS-----KMKRIISSMLDYEYVKQIK 627
QY 309 OHYOHMSDSLSEGRGLKLLKQIPIEPKKDDIHSLSOEKELLRKIQIDSSDFLSTEER 368
DB 628 EHY-HKVDITISEH-----KFOEIRQHMKDIENITHELKEMKYVQIQIDLTVY----- 675
QY 369 EFLKKIQIDIRDSLSEKELLRKIQIDSSNP-----LSEKKEFLKKLKDIOPIYN-- 422
DB 676 -----HOLENINHSLELOAQONKNIPRHNLVLEKLEITRKRRKN-KP-DISTS 722
QY 423 -----ORLODTGGLIDSPSINLDVRKQYKRDIONDALLHOSIGSTLYNKIYENMN- 475
DB 723 SHATDEQOVSOT--LIRGAHNHGDIIKGEDND---EVLILIOQSL---KTMGONQNO 773
QY 476 -----INNLATATGADLVDS---TDNTKINRGIFNEFKKNFKYSISSNYIVINER 524
DB 774 VGSILEKLNLS-----DQYQLODKLNVVEDIYKNL-RNFKIYIEKLHESKINRE 824
QY 525 PALDNERLKWRIQSLSPDTRAGYLENGKLLIQRNIGLEIKDVO-----TIKOS 571
DB 825 KFI-----TKVDVLSNYYSTLEAVYKFLHDFQEMSPFEKDELEKHLLEYEE 870
QY 572 EKEYIRIDAKVYPKSKIDPKIOEAO---LNIQEMKAKGLKRYTKYLTFNNVHNRASNI 628
DB 871 RKKYITLLEIQT--ROTLNINQNGEGDHIINNNNNN-----NANNLKKOY 914
QY 629 VSAVALLINEMKNNI-OSDLIRKVT---NYLVGNGRFEVTDITLPIAEOYTHODEYI 683
DB 915 LKDLDLEISKLGHILEVDIKKNIALEOINYLNN-----INIEYPDVIRDLMPAPRIY 968
QY 684 ---EOVHSGLYVPSRSILLHGPKSGVELRDNDSGFIHFGHADVAGYLLDKQOSDL 740
DB 969 PVSEDIYDITVWVNDNTAV-----INNTLRFVMTFPOKIDYDDHLI-----F 1012

QY 741 VTNSSKFIIDIFE-----ESSNLSYGRTNAEFFAEAF 774
DB 1013 VYNIKEL--TYKENLADKEKNYANYYEMNRFYHLHEEF 1048

RESULT 13
ID 025857 PRELIMINARY; PRT; 3119 AA.
AC 025857;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE PG377
OC Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96360472; PubMed=8719156;
RA Handman E., Osborn A.H., Symons F., van Driel R., Cappel R.;
RT "COS cell expression cloning of Pf377, a Plasmodium falciparum
gametocyte antigen associated with osmophilic bodies."
RL Mol. Biochem. Parasitol. 74:143-156(1995).
DR EMBL; I04161; AAC37257.1; -
DR InterPro: IPR001064; CrystalIn.
DR PROSITE: PS00225; CRYSTALIN_BETAGAMMA; UNKNOWN.1.
SQ SEQUENCE 3119 AA; 377358 MW; 519B99D5BDFEFC CRC64;

Query Match 5.9%; Score 235; DB 5; Length 3119;
Best Local Similarity 19.8%; Pred. No. 0.082;
Matches 174; Conservative 149; Mismatches 301; Indels 254; Gaps 43;

QY 27 FIPVQAGAGHD-----VGMHVKEKKNKDKENKRDENKTOBEHLKEIMKHIVKIE 80
DB 964 FTTISGEENHONRELKKTEANLKEEMKKRFNEQOEQRKKKAE--DENMETIOKHD 1021
QY 81 VKGEAVKKEAAKLEKVPDLEMYKAIGKITYVDG---DIRKHISLEALSD--K 134
DB 1022 METSKLEKKEEDEVTODEEFD-----KQYAIIDQOELEFLRTDSEGESDVPK 1071
QY 135 KRIK-----DIYKDALHEHYVYAKEGEYPLVYIOSSEDYV-ENTEKALNYEYI- 184
DB 1072 DKVKPPRPGSPSPFYNNATISSFHEKMELYN--TSISSLNYVKEINKRPDDVYKEIK 1129
QY 185 -----GKILSRDIISKINO-----PYQ-----KFLDVLNTIKNASDSG 218
DB 1130 SKTYPRFDLTSQTKNCNKLFGKLNETIKDKKEYOKNIOSYKKNVIDIIDIO--KKANG 1187
QY 219 QDLFTNOLKEHPTDSVEFLDONSNEVOEVPKAF----- 254
DB 1188 KYIIIONLIEIKIDYKGVNRLSDRKFYKNFKYLGKRRKMKLEDFRAQFGAIRFIKD 1247
QY 255 -AYTIEPOHROVLO-LYAPFAFNMDKFNEDQINLSLELKDQRMISR--YEKMEKIK 308
DB 1248 LTTTWEETTYVLEDIYMEKKKYKEYS-----KMKRIISSMLDYEYVKQIK 1296
QY 309 OHYOHMSDSLSEGRGLKLLKQIPIEPKKDDIHSLSOEKELLRKIQIDSSDFLSTEER 368
DB 1297 EHY-HKVDITISEH-----KFOEIRQHMKDIENITHELKEMKYVQIQIDLTVY----- 1344
QY 369 EFLKKIQIDIRDSLSEKELLRKIQIDSSNP-----LSEKKEFLKKLKDIOPIYN-- 422
DB 1345 -----HOLENINHSLELOAQONKNIPRHNLVLEKLEITRKRRKN-KP-DISTS 1391
QY 423 -----ORLODTGGLIDSPSINLDVRKQYKRDIONDALLHOSIGSTLYNKIYENMN- 475
DB 1392 SHATDEQOVSOT--LIRGAHNHGDIIKGEDND---EVLILIOQSL---KTMGONQNO 1442
QY 476 -----INNLATATGADLVDS---TDNTKINRGIFNEFKKNFKYSISSNYIVINER 524
DB 1443 VGSILEKLNLS-----DQYQLODKLNVVEDIYKNL-RNFKIYIEKLHESKINRE 1493


```
OY 525 PALNENRUKMRTIOLSPDTRAGYLNGKLLIOLRNLGLEIKYQ-----IIKOS 571
DB 1494 KFI-----TKYDVLSTNVYSTLEYWKFLLHDQEWSEFKEDELEKHLLEYEE 1539
OY 572 EKEYIRIDAKVVPKSKIDTKIOEAO-----LNIQEMNKALGLPKYTKLITTFVHNHRYASNI 628
DB 1540 RKKITTEIIOI--KDTLSTNIQNGEGDHINNANN-----NVRNNLKKOV 1583
OY 629 VESAYLILNEMKNKI--OSDLIKKYT-----NYLVDGNGRFVETDITLPNIAEQYTHODEIY 683
DB 1584 LKLDLEISKILGHILLEDIKKNIALDEQIVLYTNN-----TNEFVDPVINDLMPAPRIV 1637
OY 684 ---EYVNSKGLIYVESISILHGPSKGYELRNDESGFIHEFGHAVVDYAGIYLDKKNOSDL 740
DB 1638 PVSDIDYITWVRDNTAV-----INNTRLHFVMTFQDKIYDDHLI-----F 1681
OY 741 VTNSKKFIDIFKE---EGSNLTSGYGRTEAEFFAEAF 774
DB 1682 VYNIKEI--IKENMLADEKYNANYTEKMRPYLHLEEF 1717

RESULT 14
OY 025662 PRELIMINARY; PRT: 1939 AA.
AC 025662:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)
DE REPEAT ORGANELLAR PROTEIN.
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5825;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=96V;
RA Werner E.B., Taylor W.R., Holder A.A.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U41415; AAC63403.1;
SQ SEQUENCE 1939 AA; 229001 MW; B36E462001C6F22F CRC64;

Query Match 5.9%; Score 233.5; DB 5; Length 1939;
Best Local Similarity 19.4%; Pred. No. 0.053;
Matches 187; Conservative 170; Mismatches 280; Indels 325; Gaps 48;

OY 36 GHGDV--GMHVKEKKNKDEKRRDEKRNKTOEHLKEIM-----KH---IVKIEVKG-- 83
DB 898 GHREYVAGLEEKHKKEVVALEEKHKKEIAKLEEGH--KEVNAELGEKHKKEVVAAGLEAKHNL 956
OY 84 EAVKKEAAEKLEKVPSDV---EMTKAIGKIIYVDGITR-----HISLEALSEDKK 135
DB 957 EEGKKEVAAE--LEKRAHDVAVLEEGHKA-----EIKLGEKHKKEVVAIGIEKY 1004
OY 136 KIKDIYGDALLHGHYVAKGEYEVLYIOSEEDYVENTEKALNVEYIGKILSRDILSK 195
DB 1005 KVEAI-----KLAEEH---KD-----VYTKLGEQHKKEIALEEGHKKEVVAEVEKKNASL 1051
OY 196 INDPYQKFLDVLNTIKNASDSGDGLLFTNQLKEHPTDFSVLEFQNSNEVQYAFANAFA 255
DB 1052 LNMLEENHKNMIMIKLEEHKESASDLVEKLYQKDEYKSNKKIEELTNVIKDL----- 1105
OY 256 YVIEPQRDVLQYAPRAFNMDFNE--QRIIN--LSLELKD----- 294
DB 1106 -----NDSTIMCYKKOILIEVEKRNEXNEINIKLKYQNKMMDMKKILKEKENEIKKL 1158
OY 295 QRMLSREKMEKIKOYOH-----WSDSLSE--EGRGGLKLQIPIE 334
DB 1159 NKRTLSNKKVETKENTKNSAMVYNNENKERTIYVSVCKENISESDVEKGGCNLTKTSLK 1218
OY 335 PKKDIHLSQEEKELKRIQIDSSDFLSTEEKFLKIOI----- 376
DB 1219 -KKERNIFSINDKNE-----SSELVDTIKSAVINIKIMYKKEETEDNGCKNTEDLKNK 1269
```

```
OY 377 -----DIRDISEEKEKLLNRIOYSSNPJSEK-----EPLK-----KLKL 414
DB 1270 ILDSLNELINLENKKNVLTDENNNLKKEIEI--KKNKINKEKENNTETLLNDDIILKK 1328
OY 415 DIQYDINORLODPIGGIJDSPSIL-----DYRQYKRDIONIDALLHOSIG--STLYNK 467
DB 1329 EISEW-----KDEEEKLTKENIKLKDIEQINEKYIKKEENIMKIFENENINEVYTSLKNO 1382
OY 468 IYL-----YE-----NNNI--NLATATGADLVSDTDN--TKINGIFNE- 503
DB 1383 IEIKMKLEELKNKYELLAEKRETNMSISNDKKIYENNTLEFEDSKONNLKNVDEDKT 1442
OY 504 -----FKKNKYSTISSNMYIVDINERPALDNERLKWRIQUSPDRAGYLNGKLLIORNI 558
DB 1443 GDDINCEKNNOQAKESILKDEIKKISMLYGEEL-----NRKNSYDEKVK-----NL 1489
OY 559 GLEIKDIOI-----IKOSEK-----EYIRID 579
DB 1490 TNELKELTKRNKKEEALIAELINKLNKEKNKSVKQNDSESSNNIITRKDGKTPREYVSD 1549
OY 580 AKVVP--KSKIDTKIOEAOININQEMNKALGLPKYTKLITFVHNHRYASNIVESAYILN 637
DB 1550 DKIOCKMKANLVTLKKEP-----DLMDNINSLK-----ENFRVMSIVKE----- 1590
OY 638 EKNNNIOSDLIKKYTNLYVDNGR-----FVFPDI--TLPNIAEQYTHODEIYQVH 687
DB 1591 --NRNVQNDKIVGIVSYFKKCEKELKMDLVICLKDILSILFLNDFNLFEKIDIL 1648
OY 688 SKGLYVP--ESRSILHGPSKGYELRN---DSEGF---HEFGHAV-----D 726
DB 1649 WKQWITPEIRILFLRFSPDLKRLNRYKCVNEEYVNNERYISWALFQYILETASNLKK 1708
OY 727 DYAGYLLDKKNOSDLVTNSK-----KEIDI-----FKEGSNLT 759
DB 1709 EMIVYVLEKAKEDSCENNSSNFDKPKITDILNFSKDSIRLTKIAQLRKELNFEAKNIL 1768
OY 760 SY 761
DB 1769 NY 1770
```

```
RESULT 15
OY 0900P0 PRELIMINARY; PRT: 1786 AA.
AC 0900P0:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE LIVER STAGE ANTIGEN-3 PRECURSOR.
GN LSA-3.
OS Plasmodium falciparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5839;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RA Daubersies P., Thomas A.W., Millet P., Brahimi-Zeghidour K.;
RT "LSA-3, a conserved pre-erythrocytic malaria antigen can induce
  protection in chimpanzees.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ007010; CAB55343.1;
DR InterPro: IPR001313; PUM.
FT signal.
FT SIGNAL 1 63 POTENTIAL.
FT CHAIN 64 1786 POTENTIAL.
SQ SEQUENCE 1786 AA; 200101 MW; 5DF536D7B5B1BD98 CRC64;

Query Match 5.7%; Score 228.5; DB 5; Length 1786;
Best Local Similarity 18.9%; Pred. No. 0.078;
Matches 155; Conservative 134; Mismatches 271; Indels 259; Gaps 32;
```


THIS PAGE BLANK (USPTO)

Db 1420 IQIEKNNEESYLKRNEN--EKMSNILEEKYKL--STYEIDKNINK-IKIEDEKDE 1474
OY 128 KIKDIYGDALL--HEHYVAKGEYEPVLVIQSSSEYVENTEKALNYYEIGKILSRDL 185
Db 1475 NI-----LITKNEEINNNLEEKYKAV-----OOHLEDI-----NVLVEKOKL-----AI 1512
OY 186 SKINOPYOKFLDVNTITKNASDSODGLFTFNOLKEHPTDF--SVEFLQNSNEVQEVFA 243
Db 1513 DITREKKNITINECDKYNKMK-----LNKKLEKNONNHEHTLNITKKEQOITEREK 1566
OY 244 KAFAYIEPOHRDVLQLYAPAFNMDKFN 274
Db 1567 KNFTQKVES-----LEHAFKOSYNOLKQNE 1592

RESULT 2
O9NEU7
ID O9NEU7; PRELIMINARY; PRT: 1408 AA.
AC O9NEU7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Y39B68.M PROTEIN.
GN Y39B68.M.
OC Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitida; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Sulston J.E.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; ALI32896; CAB60918.1;
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; TYR_prot_phptase.
DR Pfam; PF002206; MSN. 1.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRTYPPHTASE.
DR SMART; SM00453; PTPC; 1.
DR SMART; SM00453; MSN; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
SQ SEQUENCE 1408 AA; 159649 MW; AC8BDD4E56A4BAA CRC64;

Query Match 9.7%; Score 135.5; DB 5; Length 1408;
Best Local Similarity 26.9%; Pred. No. 25;
Matches 66; Conservative 34; Mismatches 74; Indels 71; Gaps 12;
OY 36 VKKE--KNDENKRDDEERNKTOEHLKIMKHIYIEVGEFAVKAELKLEKVP 92
Db 909 IKKEEERKAKEREKVKKEERKAKAEERK-----VEKEKVAEEKEKVKERKKEKVK 963
OY 93 SDVLEMYAIGKIYIVDGI-TKHISLEALSEDDKKIKDIYKALLHENVAKGEYE 151
Db 964 AEE-ERLAAEEKEKVIKEKAKAEOKIKAAEEERKER--AEDA-----KVKEGOE 1013
OY 152 PVLV-----IQSEDEYVE-----NTERALNYYEIGK-----ILSRDLISKINO 190
Db 1014 KIALINAKVQEKDEKSEARRKOKEMDNAAEKILNVEKEKKEKRIIEKKIMSGDY--KDDQ 1071
OY 191 PYOKFLDVNTITKNASDSODGLFTFNOLKEHPTDFSVLEFLQNSNEVQEVFAKAYYI 250
Db 1072 RTAFKCNNTIESVASTID-----KDHITE-----SYTYLP 1102

OY 251 EPOHR 255
Db 1103 EDKRR 1107

RESULT 3
O9LM95
ID O9LM95; PRELIMINARY; PRT: 513 AA.
AC O9LM95;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE KED.
GN KED.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; eusterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LEAF;
RX PubMed=10945337;
RA Hara K., Yagi M., Koizumi N., Kusano T., Sano H.;
RT "Screening of wound responsive genes identifies an immediate-early
RT expressed gene encoding a highly charged protein in mechanically
RT wounded tobacco plants."
RL Plant Cell Physiol. 41:684-691(2000).
DR EMBL; AB009883; BAA95789.1;
SQ SEQUENCE 513 AA; 61019 MW; F9598A37A840B970 CRC64;

Query Match 9.3%; Score 130.5; DB 10; Length 513;
Best Local Similarity 28.3%; Pred. No. 14;
Matches 56; Conservative 29; Mismatches 68; Indels 45; Gaps 9;

OY 37 KEKENKDNKRDDEERN-KTOEHLKIMKHIYIEVGEFAVKAELKLEKVP 95
Db 314 KKKKKKKKKETKDEKKYKSESEEDKKEETEKKDDDEGQKRE--KEKKKKKKKK 372
OY 96 LEMVYAIIGKIYIVDGI-----ITKHISLEALSEDDKKIKND-----IYK 135
Db 373 KDKKKVKSSESDDEDDKODKVNVEVATREIKI-----EDDKKISDGADEKGRKKEKG 428
OY 136 DALHEHYVAKGEYEPVLVIQSSSEYVENTEKALNYYEIGKILSRDLISKINP----- 191
Db 429 D-----SKDEKQK-----DAKKDKAEKTRKLEDKYKSNGLKSK--LEKINAKLEAL 473
OY 192 YOKFLDVNTITKNASDSD 209
Db 474 QOKKADIMKTIKAEEDKN 491

Query Match 4
O96246
ID O96246; PRELIMINARY; PRT: 980 AA.
AC O96246;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HYPOTHETICAL 118.9 KDA PROTEIN.
GN PEB0765W.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shellen S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perlea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hofman S.L.;

RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
falciparum."
RL Science 282:1126-1132(1998).
DR EMBL: AE001417; AAC71942.1; -
KW Hypothetical protein.
SQ SEQUENCE 980 AA; 118857 MW; 72D46118BA0FF74 CRC64;

Query Match 9.2%; Score 129.5; DB 5; Length 980;
Best Local Similarity 24.2%; Pred. No. 34;
Matches 63; Conservative 47; Mismatches 95; Indels 55; Gaps 11;

QY 31 DVGMYHKEKKNKDEKRNKTOEHLKEIMKHIVKIEVGEAVAK-EAAEKLE 89
DB 743 DMEKYEEKKKNEEKKNEEKKNEEKKNEEKKNEEKKNEEKKNEEKKLE 797
QY 90 KYPSPVLENYKAIGKIYVDGDTIKHISLEALSEDKKIKIDYGDALIHEHYAKEG 149
DB 798 KKHOFEEKERK-----EIEYHOKEDRRKKDKK---KGHSDEEYKNEKEKT 844
QY 150 YEPVLVIOSEDEVVENTEKALN-----VYVEIGKILSRD---ILSKINPYQKFL-DVL 199
DB 845 KEKSSNILEDEETIIOLEERDGENCFYLL--KSLSELDVIYIKLSKDALINDAF 901
QY 200 NTIKNASDSGDLLFTNOLKEHPDTSVE-----FLEQNSNEVOEYFAKAFAYIE 251
DB 902 NKINLAITSWN---IENENKESDNTITYENTATGENTIDENTITEV-----E 946
QY 252 PQHRDVLQLYAPEAFNYMDK 271
DB 947 MNNEEYKIFSVKDYMLKK 966

RESULT 5
096195 PRELIMINARY; PRT; 1245 AA.
AC 096195;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE GAF DOMAIN PROTEIN (CYCLIC NT SIGNAL TRANSDUCT.).
GN PFB0510W.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-99021743; PubMed-9804551;
RA Gardner M.J., Tettehlin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pettersen M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser G.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
falciparum."
RL Science 282:1126-1132(1998).
DR EMBL: AE001399; AAC71891.1; -
DR InterPro: IPR003018; GAF.
DR Pfam: PF01590; GAF; 1.
DR SMART: SM00065; GAF; 1.
SQ SEQUENCE 1245 AA; 148552 MW; ED5132BED6AAD2D CRC64;

Query Match 9.2%; Score 129; DB 5; Length 1245;
Best Local Similarity 21.1%; Pred. No. 47;
Matches 62; Conservative 52; Mismatches 90; Indels 90; Gaps 12;

QY 1 KYISMSCLVATITLSPVPLVQAGGHDVGMHYKEKN--KDEKRNKDEERKTOE 58
DB 456 KLAEMKCKIISTIFMEKFTYI-----SGDDKMILTFDEKKNVKKEDDDDDYDNDND 508
QY 59 EHLKEIMKHIVKIEVGEAVAKKAAEKLLEKVPDYLENYKAIGKIYVDGDTIKHIS 118

DB 509 DEEGTELKEKEKVEEKYEDIKKKKKRQKRNILPYNIIKMEKD-DKIF-DEGD--ENRE 564
QY 119 LEALSEDKKIKD-----IY-----GKDALIHEHYAK-----EGEPLVLI 156
DB 565 SKELNEDE 624
QY 157 QSESDYVENTEKALNLYVEIGKILSRDILSKINPYQKFLDVIYNTIKNASDSGDLLFT 216
DB 625 QNGDDHDENNVELIODYVEK-----NNIYNKND----- 654
QY 217 NOLKEHPDTSVEFLEQNSNEVOEYFAKAFAYIEPQHRDVLQLYAPEAFNYMD 270
DB 655 -----TIQY-EDNNNIYKTKFSN-----NIYTKFPDNYMD 683

RESULT 6
ID 025920 PRELIMINARY; PRT; 1510 AA.
AC 025920;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE MATURE-PARASITE-INFECTED ERYTHROCYTE SURFACE ANTIGEN.
GN MESA.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-D10;
RA MEDLINE-92158014; PubMed-1741020;
RA Coppel R.;
RT "Repeat structures in a Plasmodium falciparum protein (MESA) that
RT binds human erythrocyte protein 4.1.";
RL Mol. Biochem. Parasitol. 50:335-347(1992).
DR EMBL: M69183; AAA29651.1; -
DR InterPro: IPR001623; DnaJ_N.
DR Pfam: PF00226; DnaJ; 1.
DR PROSITE: PS50076; DnaJ_2; 1.
DR SMART: SM00271; DnaJ; 1.
SQ SEQUENCE 1510 AA; 177185 MW; 95C0F57E7F6EDA9 CRC64;

Query Match 8.9%; Score 125.5; DB 5; Length 1510;
Best Local Similarity 24.7%; Pred. No. 89;
Matches 63; Conservative 43; Mismatches 110; Indels 39; Gaps 10;

QY 36 VKEKEE--KNKDEKRNKDEERKNTQOEHLKEIMKHIVKIEVGEAVAKKAAEK----- 86
DB 923 VKEKEEKEKEEVEKKEVEKESKDEKIEDEKEKEVK--EYKEDKDEKDKVIGOE 980
QY 87 -LEKVPDYLENYKAIGKIYVDGDTIKHISLEALSEDKKIKIDYGDALIHEHYVY 145
DB 981 IIEELIKKEVKRKKNNENKNDVYIOEIMNEDVNE-----KIDANRDKVIEOE--K 1033
QY 146 AKGEYEPVLVIOSEDEVVENTE--KALNLYVEIGKILSRDILSKINPYQKFLDVIYNTIKN 204
DB 1034 EKEEYKKEVEKKEVEKKEVEKKEVEKKEVEKKEVEKDESDRNVIYQ--ETIMEDVN 1090
QY 205 ASDSDGDDLLFTNOLKEHPDTSVEFLEQNSNEVOEYFAKAFAYIEPQHR-----DVL- 258
DB 1091 EKDTESDKMIGKEV-----IIEVEEYKKNVKNKVRNRRRKRKRDVIE 1140
QY 259 QLYAPEAFNYMDKN 273
DB 1141 QEIVSEEVNEKDTKN 1155

RESULT 7
ID 096224 PRELIMINARY; PRT; 558 AA.
AC 096224;
DT 01-MAY-1999 (Tremblrel. 10, Created)

Query Match 8.7%; Score 122; DB 5; Length 1624;
Best Local Similarity 25.6%; Pred. No. 1.5e+02;
Matches 58; Conservative 35; Mismatches 60; Indels 74; Gaps 12;

37 KEKKNKNDENKRDKEE-----RNKTOEHLKEIM-----KHYK---IEVKGEA 78
1053 KKQDEEKKRKEKKEEMKKNKEEKLKDIKAKSLAKESKSAKRKSLNKKDEDA 1112
79 VKKAAEKLKVPDVLNEMKKAIGKIIYVDGDTKHISLEALSEDKKIKDIYGDAL 138
1113 KKKVATD---EKVAAQARR---GTYVAMAG-----KIDTEBEAKER----- 1149
139 LHEHYVAKEGEYEPVLVIOSSSEYVENTEKALNVYEIGKILSDILSKINOPYOKELD- 197
1150 -----KE--EKVLYKKSSRLKAKNADKPKRYEVYKPY-----INDDFDKQMD 1191
198 VLNTIKNASDSGDGLFTNQLKEHPTDFSVFPL-----EONSNEVOE 240
1192 IRKQKSGS-----NOLQSAIKDKSLGKILSATEEAKSRMEE 1228

RESULT 13
006166 PRELIMINARY: PRT: 1661 AA.
AC 006166;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MATURE PARASITE-INJECTED ERYTHROCYTE SURFACE ANTIGEN (ANTIGENIC
PROTEIN PFEMP2)
OS Plasmodium falciparum.
OC Eukaryote; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_Taxid=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PALO ALTO;
RA kun J.F.J., Waller K.L., Coppel R.L.;
RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 797-850 FROM N.A.
RX MEDLINE=93122844; Pubmed=1478701;
RA Saul A., Vaganen F., Howard R.J.;
RT "Conservation of repeating structures in the PfEMP2/MESA protein of
Plasmodium falciparum";
RL Immunol. Cell Biol. 70:353-355(1992).
DR EMBL: AF056936; AAC13303.1; -
DR EMBL: S52458; AAB24869.1; -
DR InterPro: IPR001623; DnaJ_N.
DR Pfam: PF00226; DnaJ_1.
DR PROSITE: PSS0076; DnaJ_2; 1.
DR SMART: SM00271; DnaJ; 1.
KM Antigen.
KW
SQ SEQUENCE 1661 AA; 195479 MW; AF340527D85A9D29 CRC64;

Query Match 8.7%; Score 121.5; DB 5; Length 1661;
Best Local Similarity 23.8%; Pred. No. 1.6e+02;
Matches 62; Conservative 46; Mismatches 106; Indels 47; Gaps 11;

36 VKKEE--KKNKDNKRDENKKTQEHLEKIMKHIKIEVKGEEAVKKEAAK----- 86
1178 VKKEEVEKEEVEKEEVEKKTESKOKETJQEKKEEVK--EVKEDTENKQKVGIOE 1235

87 -LLKVPDVLNEMKKAIGKIIYVDGDTKHISLEALSEDKKIKDIYGDALLHEHYV 145
1236 IIEIKRKYRKYRKNKKNKNDVIOEIMNEDNE-----KDTANKDKVIOE--- 1287
146 AKEGEYEPVLVIOSSSEYVENTEKALN---VYEIGKILSDILSKINOPYOK---FLDV 198
1288 -KEKEE---VKEVEKDTVKNKDIYGOEVITIEEYKKEVKKVKNKKNKKNKNDVIOEI 1343
199 LNTIKNASDSGDGLFTNQLKEHPTDFSVFPLDONSNEVOEYAKAFAYIIEPOHR-- 255
1344 MNEDVNEKDTESKDKMGKEV-----IIEVEEYKVRKNKEVKNVNRNKNNE 1393
256 --DVL-QLYAPEAFVNDKFN 273
1394 RKDVIEQEIYSEEVNEKDTKN 1414

RESULT 14
066878 PRELIMINARY: PRT: 1156 AA.
AG 066878;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CHROMOSOME ASSEMBLY PROTEIN HOMOLOG.
G2 XCPC OR AQ_629.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_Taxid=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; Pubmed=9537320;
RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
Feldman R.A., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
"The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus";
RL Nature 392:353-358(1998).
DR EMBL: AE000699; AAC06839.1; -
DR InterPro: IPR003439; ABC_transport.
DR InterPro: IPR001687; ATP_GTP_A.
DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR InterPro: IPR002017; Spectrin.
DR Pfam: PF02483; SMC_C; 1.
DR Pfam: PF02463; SMC_N; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1156 AA; 135563 MW; B12DB30F70C0CE49 CRC64;

Query Match 8.6%; Score 120.5; DB 2; Length 1156;
Best Local Similarity 27.5%; Pred. No. 1.2e+02;
Matches 64; Conservative 31; Mismatches 89; Indels 49; Gaps 11;

31 DVGHNKVEKKNKDNKR-KDEERNK-----TQEHLEKIMKHIKIEVKGEEAVKKEAA 84
340 EVGTLQLELEKLEKYEKSLKEVEREKRLLEEEERKLTIFDEYKLE-----EEK 390
85 EKLEKVP-----DLEMYKA-IGKIIYVDGDTKHISLEALSEDKKIKDIYGDAL 138
391 EKLTKLNSLNKKEQELIORANKTKIERIKEDINK-----LISREKIKETKEKEOE 445
139 LHEHYVAKEGEYEPVLVIOSSSEYVENTEKALN---YEIGKILSDILSK--INOPY 192
446 IKRLAIKAKK-----EEELRNLTQELNITKRLSVKRRKLEVEKKAIRREV 495
193 OKFLDVLNTIKNASDSGO--DLFTNQLKEHPTDFSV-----EFLDONSNEV 238
496 RSFSDVSVDFKDIKGVYGSVELIRVKNPEHITAIEVAGGRLKFIIVEDEEV 548

THIS PAGE BLANK (USPTO)

QY 181 YVEIGKILSRDILSKINOPYOKFLDVLNTIKNASDSOGDILFTNOLEKHPDSEVFLE 240
DB 181 YVEIGKILSRDILSKINOPYOKFLDVLNTIKNASDSOGDILFTNOLEKHPDSEVFLE 240
QY 241 QNSNEVOEFKAFAYIIEPOHRDVLQYAPAFNMDKFNQEOEINLSLEELKQRMLSR 300
DB 241 QNSNEVOEFKAFAYIIEPOHRDVLQYAPAFNMDKFNQEOEINLSLEELKQRMLSR 300
QY 301 YEKMEKIKOHQWMSDSLSEEGRLKKLQPIEPKKDDIHSLSQEEKLLKRIQIDSS 360
DB 301 YEKMEKIKOHQWMSDSLSEEGRLKKLQPIEPKKDDIHSLSQEEKLLKRIQIDSS 360
QY 361 DFLSTEEKEFLKQIQRDLSLSEEEKELLNRIOVDSSNPLSEKEKEFLKRLKIDQPYD 420
DB 361 DFLSTEEKEFLKQIQRDLSLSEEEKELLNRIOVDSSNPLSEKEKEFLKRLKIDQPYD 420
QY 421 INQRLQDTGGLIDSPSLNLDVRKQYKRDIONIDALLHOSIGSTLYNKYYELNMINNLT 480
DB 421 INQRLQDTGGLIDSPSLNLDVRKQYKRDIONIDALLHOSIGSTLYNKYYELNMINNLT 480
QY 481 ATLQADLVSTDNTRKINGKIFNEFKKNEKYSISSNMYIVDINERPALDNERLKWRIDLS 540
DB 481 ATLQADLVSTDNTRKINGKIFNEFKKNEKYSISSNMYIVDINERPALDNERLKWRIDLS 540
QY 541 DFRAGYIENGKLLIQRNIGLEIKRDOYIIKOSEKEXIRDAVVRKSKIDPTIOAOQIN 600
DB 541 DFRAGYIENGKLLIQRNIGLEIKRDOYIIKOSEKEXIRDAVVRKSKIDPTIOAOQIN 600
QY 601 QEMKALGLPKYTKLITFNHNRYSNIVESAYLILNEMKNNISODLIKVTNYLVGNG 660
DB 601 QEMKALGLPKYTKLITFNHNRYSNIVESAYLILNEMKNNISODLIKVTNYLVGNG 660
QY 661 RFVETDITLPIAIOYRHODEIYEQVHSGKLYVPSRSILHGPSKGYELNDSGEFHE 720
DB 661 RFVETDITLPIAIOYRHODEIYEQVHSGKLYVPSRSILHGPSKGYELNDSGEFHE 720
QY 721 FGAHVADYAGLILDKNOSDLVTNSKKFFIDIFKEGNSLTSYGRTEAEFFAEFLRH 778
DB 721 FGAHVADYAGLILDKNOSDLVTNSKKFFIDIFKEGNSLTSYGRTEAEFFAEFLRH 778
RESULT 2
B59106
hypothetical protein pXOI-122 - Bacillus anthracis virulence plasmid pXOI
C:Species: Bacillus anthracis
C:Date: 12-Nov-1999 #sequence, revision 12-Nov-1999 #text_change 09-Jun-2000
C:Accession: B59106
R:Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Kelm, P.; Koehler
J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of pXOI, the large Bacillus anthracis plasmid harbored
A:Reference number: A59091; MUID:99445483
A:Accession: B59106
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-800 <OKT>
A:Cross-references: GB:AF065404; NID:g4894216; PIDN:AD32426.1; PID:g4894338
A:Experimental source: strain Sterne
A:Note: Similar to calmodulin sensitive adenylate cyclase, edema factor, cya, plasmid pX
C:Genetics:
A:Gene: pXOI-122
C:Superfamily: calmodulin-sensitive adenylate cyclase; calmodulin-sensitive adenylate cy
P:34-286/Domain: lethal factor amino-terminal homology <LFA>

Query Match 12.6%; Score 501; DB 2; Length 800;
Best Local Similarity 24.6%; Pred. No. 3,6e-13;
Matches 202; Conservative 145; Mismatches 281; Indels 194; Gaps 36;

QY 4 KKEFI-----KYVMSCLVATITLSGPFPIPLVGAGGHDVGMVKEKKNKDKRDE 59
DB 3 RNRKIPNKFISPSVLLFAISSQAEVNA-----MNEHYTSSDIKRNHKT 49

QY 60 ERNKTQEHRLKEIMKHIVKIVGGEAVKKAERKLEKPSVDLEMKATIGKITIYVDC 119
DB 50 EKNTEKEKFKDSINNLYVKEFTETLDKIQOTDILLKIPKVDLEYSELSEGEIYFTDI 109
QY 120 DITHSILEASEDEKPKKIKDIDYGKDALHEHYVAKEGEYFVLVIOSESDVENTKALN 179
DB 110 DIVEHKELODLSSEEEKSMNSRGEKVPFASRFVEKREPKLII-NIKDVAINSQESKE 168
QY 180 YVEIGKILSRDILSKINOPYOKFLD--VLNTIKNAS--DSOGDILFTNOLEK---HPT 232
DB 169 YVEIGKILSRDILSKINOPYOKFLD--VLNTIKNAS--DSOGDILFTNOLEK---HPT 232
QY 233 DFLSTEEKEFLKQIQRDLSLSEEEKELLNRIOVDSSNPLSEKEKEFLKRLKIDQPYD 292
DB 224 SIDINFIKENITFEOHAFSLAFSYFPDHRVLELYAPDMFEYENKLU----- 271
QY 293 KQOYMLSRYEK--WEKIKOHQWMSDSLSEEGRLKKLQPIEPKKDDIHSLSQEEK 350
DB 272 -----EKGEFKI-----SESLKKG-----VKDRIDVL---KGEKA 301
QY 351 LKRIQIDSSDLSLSEEEKFLKQIQRDLSLSEEEKELLNRIOVDSSNPLSEKEKEFLK 410
DB 302 L-----KAGLVBEHADAFKI-----ARLNTYILFRPNKLTATN---LI 339
QY 411 KLRKIDQPYDINQRLQDTGGLIDSPSLNLDVRKQYKRDIONIDALLHOSIGSTLYNK 461
DB 340 KSGVATKGLNVHGKSSDGPVAGYIPDQDLSKRGQOLAVENKLENKKSITHEGEITG 399
QY 462 STLYNKIYL--YENMINNLTATLGADLVSTDNTRKINGKIFNEFKKNEK-----YSISSN 515
DB 400 -----KIPKLDHLRIEELK-----ENGILKCKKEIDNCKKYYLLESN 438
QY 516 YMIVDINERPALDNERLKWRIDLSLSEEEKELLNRIOVDSSNPLSEKEKEFLKRLKIDQ 573
DB 439 NOVVEF--RISDENNEVOYK-----TKEGKITVGEFFNMRNIEVMAKNVEGKLPYLA 490
QY 574 EYIPIADKVVKSKIDPRKIOEAOINQEMKALGLPKYTKLITFNHNRYSNIVESAYLILN 625
DB 491 DY---DLFALPSSLTEIKKQIPQ---KEMDKVYNTPSLEKQKGYNLTLYKYEIERK-- 541
QY 626 SNIVESAYLILNEMKNNISODLIKVTNYLVGNGREYFTDITLPIAIOYTHOD--EIV 683
DB 542 ---PDSTKGLTSMNQKQMLDRL-----NEAVKYTGTYGADVNVHTEGDNEBFP 587
QY 684 EQVHSGKLYVPSRSILHGPSKGYELNDSGEFHEGHAVDYA---GYL--- 732
DB 588 EKDNIEFLINEGEFIILTKMNMETGRFEKNTKRDVLYFNRSYKNIAPGNKAYIEWTD 647
QY 733 -LDKNOSDLVTNSKKFFIDIFKEGNSLTSYGRTEAEFFAEFLRH 773
DB 648 PITKAKINTIPTSAEFL-----KNLSSIRSSNVGVYKDS 682
RESULT 3
J50029
adenylate cyclase (EC 4.6.1.1) precursor, calmodulin-sensitive - Bacillus anthracis
A:Alternate names: anthrax toxin edema factor
C:Species: Bacillus anthracis
C:Date: 31-Mar-1992 #sequence, revision 11-Nov-1994 #text_change 19-Jan-2001
C:Accession: J50029; PS0307; J50602
R:Robertson, D.L.; Tippetts, M.T.; Leppla, S.H.
Gene 73, 363-371, 1988
A:Title: Nucleotide sequence of the Bacillus anthracis edema factor gene (cya): a cal
modulin-sensitive adenylate cyclase
A:Reference number: J50029; MUID:89211974
A:Accession: J50029
A:Molecule type: DNA
A:Residues: 1-800 <ROB>
A:Cross-references: GB:M24074; NID:g142812; PIDN:AAA79215.1; PID:g142813
A:Accession: PS0307
A:Molecule type: protein
A:Residues: 34-48 <RO2>
R:Escuyer, V.; Duflo, E.; Sezer, O.; Danchin, A.; Mock, M.
Gene 71, 293-298, 1988

QY 350 -----ELKKRIQSDSFLSTE---EKFFIKKQIQRIDRISSEKKL-----Ln 391
Db 1650 CMLFTHHTLAEFLKIKIDYDYSKFIESATKFSKEFLAKYIG-DTSNSLNDIATLQKLYDH 1708
QY 352 RIQ-----VDSNPLSEKEKEFLKLLKLDIQPYDINOQLDGTGGLDPSINDV 441
Db 1709 QINKVYTSKLSDPATNDNNLLEKE-----ATQAIKNLTKLFTIDSNNDIA 1755
QY 442 RKQYKRDIONI--DALLHOSIGS--TLNKKIYLENMI----- 476
Db 1756 NALHNKKIOMYFENSELHSIESIKQLYKKMHVFKLLINGQINGYEDISKOFDNILOL 1815
QY 477 -NNLTATLGLADL--VDSYDNTKINRGI-----ENEKKNKFSYSSNYMIVDNE 523
Db 1816 ESELFLANLN-DLKEIGQIKSDKKNFHLALNETHPIPNFTLKEIYHDLYKKRQIDELEN 1874
QY 524 RPDALNE-----RLKMRIOQLSPDPRAGYLENGKLLIQRNIGLEIKDVOIIGSEK 573
Db 1875 ITSESENNITLYIDITTKLKEKQVOSILNFVPTYENDSNIKOHIDOTENDENSKIKESL 1934
QY 574 EYIRIDAVKVPK-----SKIDRKIQEAOQINQEW--NKALGL 609
Db 1935 TTIQSFQELINKINGIKAKOFDNNNNINNNNTITSSODVDVKKHHSKDLTTELELEI 1994
QY 610 PKYTKLI---PFNV---HNRYASNIVESAYLLINEMKNNIQ---SDLIKVTNLYVGN 659
Db 1995 OKSLIEDINKSTYEINGANNVNTYIRNYVEQOTNKIQNNSKNDELIDLIQKILNKRK-- 2052
QY 660 GRFVTDITLPIPAQYTHODETYQVHSGKLYPESHSILHPSKCV-----ELRND 713
Db 2053 -----SEKRLPITGNKKNNVTSIISRI-NKYVNLIESE---YGNNNNVSYVAKKEED 2102
QY 714 SEGFTHETGHAVDVAAGYLLDKNOS--DLVYNSKKPDIRK-----EEGSNLTYSERT 764
Db 2103 ANSLILD-----LDKSNHILKLDLOQNLKIITDKKKNKOEIENRNNOITNRE 2150
QY 765 NE 766
Db 2151 QE 2152

RESULT 5
C71622
Hypotheetical protein PF00145c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: C71622
R:Gardner, M.J., Tetteh, H., Carucci, D.J., Cummings, L.M., Aravind, L., Koonin, E.V.,
Perle, M., Salzberg, S., Zhou, L., Sutton, G.G., Clayton, R., White, O., Smith, H.O.
Science 282, 1126-1133, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743
A:Accession: C71622
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1979 <GAR>
A:Cross-references: GB:AE001375; GB:AE001362; NID:g3845105; PIDN:AACT71819.1; PID:g3845101
A:Experimental source: clone 3D7
A:Genetics:
A:Gene: PF00145c

[illegible]

A:Cross-references: GB:M34281

Query Match 6.3%; Score 253; DB 2; Length 2401;
Best Local Similarity 20.2%; Pred. No. 0.012;
Matches 182; Conservative 159; Mismatches 340; Indels 222; Gaps 38;

```
OY 2 NIKKEPIKYSCLVTAITLSPVPFIPVVGAGHGVDGMHKEKKKDEKRRDEER 61
|||
DB 1389 NIYEPIKSYDLITHYLEVSEKPIYEDIKRRTAQNELTNKKNVAKASYLDDIER 1448
|||
OY 62 NKTQE--EHLKEIMKH-----VKIEVGEE-----AVKKEPAKLEKVPDYL 104
|||
DB 1449 NEFDRIYTHFKKLVNDVNDKFTNEYSKVN--KGFDNISNSINNVKSTDEMLNLINQK 1507
|||
OY 105 EMYKAIGKIYI-----VDG-DITRHS---LEALSEDK 134
|||
DB 1508 EMYANIVSKRYYSYKAEENIFINIPKLANSLNIQKSSGIDLFKNINAIPLYLDSQK 1567
|||
OY 135 KIKIDYGDALLHEHYVAKEGEYEPVL-VIOSEEDYVENTEKALNVYEIGKILSRDL 193
|||
DB 1568 KOTLPIFPEPKTSETTYTISDYNTLIDILKRSOELQKEDQALNLIFE-----NRLH 1622
|||
OY 194 SKINOPYOKFLDVLNTIKNASDSGDGLFTNQLKEHPTDFSVFELEONSNEVOEFAPA 253
|||
DB 1623 KRY-QATNELKDTLSDLKRRKEO-----ILNKVK-----LLHKSNEMLKLSGNS 1666
|||
OY 254 FAY--YIEQHDVQLVLAPEAFNMDKFNBOEINLSLELDQRLSREKREKIKOH 311
|||
DB 1667 QNYDTILESSKYDKIK--EKSNNYEKEENLGINFDVYAMEBO--FNNDIKDEKLENNY 1722
|||
OY 312 OHMSD---SLSEGRGL---KKLO-----PIEPK--KDDIHSLOEKE-- 350
|||
DB 1723 KISEKDNVFPSENNNIILOSKKKLKELTANFAEIKEDKIEKGLINKLETREKDCM 1782
|||
OY 351 -----LAKRIQIDSSDF-----LSTEEKFLKQIDIRDSLSEEEKEL-----LNRI 393
|||
DB 1783 LFTYKLVETLKIKTQDYKFTSATKFSKEFLKYIDA--TNSLNDINDITLQKRYDLNOI 1841
|||
OY 394 Q-----VDSNPLSEKEKEFLKKLQIDPIYDINORLODGTGGLDPSINLDYRK 443
|||
DB 1842 NKHVASVADATNDNNNLIEKEKEATKTINLTFL-----TIDSKRIDADGLH 1890
|||
OY 444 OYKRIQIONIDALLHOSIGS--TLVYKITYLENNMI-----NNTLATGADL 487
|||
DB 1891 NKKIQIYNSLHKSIDSIKQIYKMAFKLLINIGHIKKYPFDISKEDNIIQLOESEL 1950
|||
OY 488 VSDNTKINRGIFNEKKNFKYSISSNYMIVDINERPALDNERLKWRIQLSPTBAGYL 547
|||
DB 1951 TANLNDLKEIGOKISDKKQFLHALSET--PIPNFNLKEIYHDIYKYNQOIDEINITNE 2009
|||
OY 546 ENGKILLOINIGLEI-----KVOIHKOSEKEIRIRDAVVPKSKIDTKI 592
|||
DB 2010 EENNTITLYDITLTKLKKVOSILNEVTYENDSNIKHQIIONNENDVSKI--RESLETTI 2068
|||
OY 593 QEAQLINQEMKALGLPYTKLITFNHNRVANSIVESAUYLLNEMKNNOISQDLKRYT 652
|||
DB 2069 OSFOKRLN-----KLNEIKAFYDNNNNINNVISITISQDVNDVKKHISDL--TIE 2116
|||
OY 653 NTLVDNGRFVFTDITLPLNIAEOY--THODEIYEOVHSGLYVPESRSILLHGSKVEL 710
|||
DB 2117 NELLIQ-----IQLSLEDIKKSTYDIRSEQITKYVNPJHDVYEQDTKKIQQNPNK---- 2165
|||
OY 711 RNDSEGFHEFGHADVAGYLLDKN-QSDL-----VTNSK-----KRIDFKBE 754
|||
DB 2166 -----DEIDDLIQEIVNYSKESLKLPTIINKNDVTPITISRIDKVINILKSK 2213
|||
OY 755 GSN 757
|||
DB 2214 YNN 2216
|||
```

RESULT 7

T28317
orf MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C:Species: Melanoplus sanguinipes entomopoxvirus
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T28317
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J:Virology 73, 533-552, 1999
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: 220484; MID:99102612
A:Accession: T28317
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1127 <AFO>
A:Cross-references: EMBL:AF063866; NID:g4049647; PID:AMC97677.1; PID:g4049717
C:Genetics:
A:Note: MSV156

Query Match 6.0%; Score 239; DB 2; Length 1127;
Best Local Similarity 20.8%; Pred. No. 0.017;
Matches 159; Conservative 121; Mismatches 252; Indels 234; Gaps 35;

```
OY 60 ERNKTQEE--HLKEIMKHIVKIEVGGEAVKKEAEKLEKVPDYLEKMYKAGIKYI- 115
|||
DB 24 ENKVSILDIINSLEYELNNI-----KFSDKITNELIKYKNIYERKIFY 66
|||
OY 116 -----YVGDITKHSLEALSEDKKKID-----YGRDALLHEHYVAKEGEYEP 161
|||
DB 67 MHNQFINDYNTILOYLEENNEINCKIKENKPPCKNPLY-----NITYKKKLYID 118
|||
OY 162 LVIOSEEDYVENTEKALNVYE-----IGKILSDILSKINOPY-----OKFLDYL 207
|||
DB 119 L-----DIEKKDKELVINIEQKNAVDKI--NDIKNNVNIHSDNETIITGETLIDL 170
|||
OY 208 NTKNASDSGDGLFTNQLKEHPTDFSVFELEONSNEVOEFAPAYIIEPQHDVLO 267
|||
DB 171 NKLKLVSSDEKOLI--EQIYKNNKKEIF--KNIDVQKEIK----- 211
|||
OY 268 LVAPAFNMDKFNBOEINLSLELDQRLSREKREKIKOHYOHMSDSLSEEGGLK 327
|||
DB 212 -----KQELINKLDE-----SKREFK----- 229
|||
OY 328 KLOPIEPKDDIHSLOEKEELKRIQIDSSDFLSTEEKFLKQIDIRDSLSEEEK 387
|||
DB 230 -----KQELINKTIDKQEBELIKKNDKEINFPNDEKQKLLDQINSKI--NTLWENIK 280
|||
OY 388 ELINRIQDVSSNPLSEKEKEFLKKLQIDQYDINORLODGTGLIDPSINLDYRKQYKR 447
|||
DB 281 GYVNI--LYTEKKNKISNLONEILNK--DSTIKSLDEKQKILD-----ELDK 322
|||
OY 448 DIONIDALLHOSIGTLVYKITYLENNMNTNLTATGADLVSDNTKINRGIFNEKKN 507
|||
DB 323 NINNTI-----TSLYNN-----SNKTTNIQOULLESSLTD--FNANNTI--INELSK 365
|||
OY 508 FKYSISSNYMIVDINERPALDNERLKWRIQLSPTBAGYLENGKLLLORNIGLEIRVOI 567
|||
DB 366 IK-----LFD--NLOQKLNNDITEQNNKIT-----DFPNNSTRIFKEKLDIYEVKIDD 411
|||
OY 568 IKOSEKEYIRIDAKVVPKSIDTKIQEAQLINQEMKALGLPKRYKLTENVHNR---- 623
|||
DB 412 IKNNNLOKLEESYK-----KIDQTEYKKNKINKEVNDITEL--KNNNLOKLEENKIDE 465
|||
OY 624 ---YASNIYESAULLINEMK--NNIQ-----SLLIKVTYVYLDGNRFPFTDITLP 670
|||
DB 466 QTEYKRNKINKEVNDITELKNNNLOKLEENKKNINOKLTYLKNDISNTELF-----NKL 520
|||
OY 671 NIAEQYTHODEI-----EYOVHSGLYVPESRSILLHGSKVELRNDSEGFIFHEGHA 724
|||
DB 521 NISDPKDKSREIAKLNTEYQL-----RKDLLENINKTNELMLKLSDNKLSLEQL 570
|||
OY 725 VDDYAGYL--LDKNQSDLVYNSKKFIDIFKEGGSNLTSGRTNEAF 768
|||
DB 571 YDSKKNILDGIDIKIYNSLAKKNDKIDEIF-----SNIEKFDIYVIE 612
|||
```

[illegible]

```

D0 1584 LKDDLETSKLGHILEVDIKKNALAEQINVLNN-----INETVPDVIRODMAPRIV 1637
Q0 684 ---EOVHSGGLYVPSRSITLLHGPGKGVLRDSEGFIEFGHAVDYAGYLDDKQSD 740
D0 1638 PVSEDIYDITTVVRDNTAV-----INNTLRHFVMTFDOKIVDYDHLI-----F 1681
Q0 741 VTNKKFDIFKE---EGSNLTSGRTNEAEFFAAAF 774
D0 1682 VYNIKEL--IYKENLADEKYNANYEMRFLHLEEF 1717

RESULT
9
118372
Repeat organelle protein - Plasmodium chabaudi
G:Species: Plasmodium chabaudi
G:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
G:Accession: T18372
R:Werner, E.B.; Taylor, W.R.; Holder, A.A.
Mol. Biochem. Parasitol. 94, 185-196, 1998
A>Title: A Plasmodium chabaudi protein contains a repetitive region with a predicted
A:Reference number: Z18922; MUID:98418765
A:Accession: T18372
A>Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1939 <WR>
A:Cross-references: EMBL:U03145; NID:g1151157; PID:g1151158; PIDN:A063403.1

Query Match 5.9%; Score 233.5; DB 2: Length 1939;
Best Local Similarity 19.4%; Pred. No. 0.05;
Matches 187; Conservative 170; Mismatches 280; Indels 325; Gaps 48;

Q0 36 GHGDV--GMHVEKEKNKDKNRKDEERNKTOEHLKEIM---KH---IYKIEYKG-- 83
D0 898 GHREYVAGLEEKKKEVVALEEKHKKEIKLEEGH--KEYVAELGEKKKEVVALEAKHNL 956
Q0 84 EEAIVKKEAEKLEKVPSTVL---EMYKAIGIKIIVDGDTL---HLSLEALSDKK 135
D0 957 EEGHKEMVAE--LEKRHADLVAVLEQHKH-----ETIKLGEHKKEVVAIGIEEKI 1004
Q0 136 KIKDIYGDALDLHEHYVAAKEGEPELVYIQQSSSEVDYENTEKALNLYVEIGIKILSRDILSK 195
D0 1005 KVEAL-----KLAEEH---KD-----VYTKLGQKHKEIAKLIEDGHKEVYANEVKEKNASL 1051
Q0 196 INQPIQKELDVNTITKNASDSGQDILFTNOLKENHTDSEVPELEQNSNEVDGVFKAFA 255
D0 1052 LNMLEENHNKNEIKLKEHKESASDLVEKLYQKDEEVKSNMKIEELTIVYIKDL----- 1105
Q0 256 YVIEPQHHRDVLQVAPAFNYMDKFE--QETN---LSLEELKD----- 294
D0 1106 -----NDSTICYKQILIEVEYKRENYEELIKLIVONEKMDKMDKILKEKENETKL 1158
Q0 295 QRMLSRYEKWEKIKQYOH-----WSDSLSE--EGRGLLKKLIQIPIE 334
D0 1159 NKKLSNYKVFETKEKNTYKSEMYVNEKRPRIIVDSYCKENISSEVDGKGNLKMTLISL 1218
Q0 335 PKKDDIILSLQSEKELLKRIQIDSDFLSTBEKELKLQI----- 376
D0 1219 -KKEKNIFISINDKNE-----SELVDITKSALINKINEKKIEKEDNGKNIEDLKN 1269
Q0 377 -----DIDSLSEEEKELNRIQVQSSNPLSEK-----EFLK-----KLKL 414
D0 1270 ILDLSELNILENMKNVNLDENNNLKKEIET--KDNKLNEKKNEMNELINDIILKLLK 1328
Q0 415 DIQPIQDINORLODTGGLDPSINL-----DYAKQYKRIQIINDALLHOSIG--STLYNK 467
D0 1329 EISEW-----KDEBKILKENIKLNDIEQINKEKRIEEMLMIKFENINEINETSILKQ 1382
Q0 468 IYL-----YE-----NMN--NNLTATIGADIVDSTQN--TKIRGFGFE-- 503
D0 1383 IEIEKMKLEELNKNTELLAEKRETMGISNDNDKATYVENNILEDTDSKONNLKNKNEVDT 1442
Q0 504 ----FKKNFKYSISSNMYIVDINERPALDNERLAKWRIQLSPDTRAGYLENGKLILQIRNI 558

```



```
Db 1443 GDDINCKNNDDAKKEISYLDKDEIKKISMLYGEEL-----NRKNKYDEKVK-----NL 1489
QY 559 GLEIKDVOI-----IKOSEK-----EYTRID 579
Db 1490 TNLKELKINKKKEEFAELNKLNIKENKNKSVKONDESSNNITTKDDKDPPEYVSD 1549
QY 580 AKVVP--KSKIDTKIOEAQNLINOENMKALGLPKYTKLTFNVHNRASVIVESAVILN 637
Db 1550 DKIOKDKKALVYLKLEKP-----DLMDNINSLEK-----ENFRVMSIVE----- 1590
QY 638 EMKNNTIOSDLIKKVTNYLVGNGR-----FVFTDI--TLPNIAEQYTHODEIYQVH 687
Db 1591 --NKVNOQNDKIYGIYSYFKKCEKELKNDMLVICLVKLDLISLEFLDNFPYNLEKIDKIL 1648
QY 688 SKGLYVP--ESRSTLHGPSKGVLELRN---DSGCFI---HEFGHAV-----D 726
Db 1649 WKQMYIPTETIRIFLRFYFSFLDKLRNVKCVNEEYVNNERYEYSWALFQYLETASMLK 1708
QY 727 DYAGYLLDKKQSDLYWNSK-----KRFIDI-----FKREESNLT 759
Db 1709 EMYYVLEKRAEKDSCENSSNFPDKPTITDILNFSKDSIRLKTIAQLKELNFEREAKNIL 1768
QY 760 SY 761
Db 1769 NY 1770

RESULT 10
A42771
reticulocyte-binding protein 1 - Plasmodium vivax
C:Species: Plasmodium vivax
C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 28-Apr-1995
C:Accession: A42771
R:Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
Cell 69, 1213-1226, 1992
A:Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.
A:Reference number: A42771; MUID:92315338
A:Accession: A42771
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2829 <GAL>
A:Experimental source: Belem strain, merozoites
A>Note: sequence extracted from NCBI backbone (NCBIN:108114, NCBIP:108115)

Query Match 5.8%; Score 232; DB 2; Length 2829;
Best Local Similarity 18.8%; Pred. No. 0.1;
Matches 190; Conservative 159; Mismatches 315; Indels 346; Gaps 46;

QY 39 DVGMMHKEKEKNKNDENK-----KDEERKKTQDEHKEIMKHIVKT----- 79
Db 706 DINALLEEVEKVTENKESTLEMIKDEE---MEEKLDAKETFAKLNFEVSDKLTIDVYT 761
QY 80 ----EYKGEAVKKEAEKILKEKVPDLYBMYKAIGKI-----YIVGCD--ITKHI 125
Db 762 KMSAEVTNMEGIKKELAQKQFENVHKKMKFSDAFSTKFEALQNSMQQYNOEGDAIKKH 821
QY 126 SLEALSDEK---KKIKDIYKQALLLHEHYVAKEGY---EPVLVIQSSSEDVVENTEK 176
Db 822 QNRSEKEEYFKNESVEEDLSRETEOEYTKKHNNFSRRKSGISAITNMREVINKIES 881
QY 177 ALNVVYEI-----CKILSRDILSKINOPYOKRLDVNLNTIKNA--- 213
Db 882 QLVYGVIEKYFSLIGDQNEVSTAKALKETIVSDLSLKDIDQYETERKEKTSAVENTVST 941
QY 214 --SDSDGODLL-----FTNOLKEHPTDES----- 235
Db 942 IQSLSKAISLKLKLNKINSCKKYNDIDILRSKIKTLREEVOKEMPKRDKGCGENTTAL 1001
QY 236 -VEFLLEONSNEVOEVPKAKAFAYIEPQHRDVLQLYAPAFANVMDKFNQEOINLSLEELKD 294
Db 1002 LKSLSLDKMKGINELKLDGRSLNLDTFKKEDLLKLFYSE-----SKSKIHLSKDKGP 1052
```

```
QY 295 QRMLSREKKEKIKO-----HYQHWSDLSSEGRGLLKLOI-----PIEPKKDDIHS 343
Db 1053 QDPLNRIDEMEDIKRVDLNVNTO---VISEKVTLPFNKNSVYIEAMSHINTVAHG 1108
QY 344 LSOEKEKLLRIQIDSSDFLSTEEKLEFKL-----QID--INDLSSEKELNLR--- 392
Db 1109 ITSCKNKLILSVKKEVEKLNLYEONEDYKVKYKNPENKOLEAIGSMS-KLVEYINKHVS 1167
QY 393 --IQVDS--SNPLSEKER-----EFLKKYLKIDQPYDINORLOD 427
Db 1168 EMTOLESTANTLKSNAGKRENEHDLFEILNTKQGMRDYIEKTKIABELEKGYANE-LKD 1226
QY 428 T-----GGIDPSIMLD-----VRKQKIDIONI-----DAL 455
Db 1227 ANEKANKVEPEPERNIIGHLEIRITVEKDKAGKVEEMNSLTKIEKLTIEDTSDSQNEL 1286
QY 456 LHOISGSLYN-KLY-----LYENMNINLTATLGA 485
Db 1287 VTTISITHLENAKYEDVIRKNEEDSIOLEKAKSLETLDKMKLYVOOVNMLNLSAIOG- 1345
QY 486 DLYDSTNTKINRGI---FNEFKNFRYSISSNMYIVDINERPALDNERLKW-RIQLSPD 541
Db 1346 -----NAGISKELNELKGVIELLISTNYSI-----LEVYKKNSSSES 1382
QY 542 TRAGYLENGKIL-----QRNIGLEIKDVOIIOK---SEKEYIRIDAKVVPKSKID----- 589
Db 1383 VRESOLANGEFYTAKEGEEKNASARLAEERKEKEDIVKLDSDIDDV---KKEGKIKRE 1439
QY 590 -TKIOEAQLINDEMNKALGLPKYTKLTFEVNH-----RYASNIVESAVYILNEMK 640
Db 1440 ILMKKESALTF--WEES-----EKFKQKSSIMEAKGKKKIELKNNGDCG-----K 1486
QY 641 NNIOSDLIKKVTNYLVGNGRFVFTDITL-----PNIAEQYTHODEIYQVHSGLYV 693
Db 1487 ANITDSOMEVEGVNSKAEHAFHTEVAQVDKTKAFCESIYAVYTKMNLFNESIMKEVKY 1546
QY 694 PESISILLHGPSKGVLELRNDESGFIHEFGHAVDYAGYL-----LD 734
Db 1547 K-----CEKKNDE---AEKYSALKPYDGRITARVSENERKISSELKAKVE 1590
QY 735 KNOS----DLYTNSKKRFIDIFKEE--GSNLTSYGRT--NEAEFFFAELRLM 777
Db 1591 KKESQQLNDVSTKSLDIDNCRQQLDVLNIGGVKQNALQYFDSAPKSM 1640

RESULT 11
S28061
SCP1 protein - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 08-Oct-1999
C:Accession: S28061
R:Meuwissen, R.L.J.; Offenberger, H.H.; Dietrich, A.J.J.; Riesewijk, A.; van Iersel, M.
EMBO J. 11, 5091-5100, 1992
A:Title: A coiled-coil related protein specific for synapsed regions of meiotic proph
A:Reference number: S28061; MUID:93099884
A:Accession: S28061
A:Molecule type: mRNA
A:Residues: 1-946 <MEU>
A:Cross-references: EMBL:X67805; NID:957212; PIDN:CAA48006.1; PID:957213
A:Gene: SCP1
A:Keywords: DNA binding

Query Match 5.8%; Score 229.5; DB 2; Length 946;
Best Local Similarity 21.8%; Pred. No. 0.032;
Matches 192; Conservative 154; Mismatches 331; Indels 205; Gaps 48;

QY 44 VKEKKNKDKENKRRDEERNTQDEHKEIMKHIVKIE-----VKGEAVK----- 88
Db 77 LKQKENKIQENRKTIEAQRAIQELQFENKRVSLKEEIEIQENKDLIKENNAATRRHMCNLL 136
```

QY 89 KEAAKLEKPSVPLEYKAIKGIYI -VDGDTIHI -----SLEKLSDKK-----KIK 138
 137 KETCARSAEK--TSKYEYEREPFQYVVDLNNIEKMIILAFAELKRAQANALNHEHFKLK 194
 QY 139 DYGKDALLHEHY---YAAKEGEPVLVIOSSEDEVYVENTERKALNYYEIGKILSDILSK 195
 Db 195 EDHEHIOHLEEEYOKREVNKNENOVSLILOSTER - ENMKKIDTFLLE-----ESRD--K 245
 QY 196 INOPYOKF-----LDVLTNTKNAISDSDGDDLLFTNO-----LKEH-----PTQFSVEF 238
 Db 246 ANOLEKREKLODDENIKELNEKHDLTSELEDIKMSQGRSMSTOKTLEBOLQATATTIYQL 305
 QY 239 LEONSNEOEY-----FAKFAAYIE-----POHRDVLQIAYE----- 272
 Db 306 TEKKACQOEELINKAKTTHSLVLTLEKATCTTLEELRTTEQORLENNEODLKITTELQKK 365
 QY 273 --AFNYMKF--NEAEINLSLEELK-----DORMLSYEKMEKIKQIYHWSNLS--EE 321
 Db 366 SSELMEKFKFNKNKEV--ELBELTTLIAEDOKLDEKQOYKAELELOKDEOFELFLIQT 423
 QY 332 GNGLLKQIPIEBRKDDIHSLSQE-----EKELLRIQID--SSDPLSTEEKF-- 370
 Db 424 REKEIHLEVOVTVTKTSEEHYLQVEMTELEKELKNIETLANSDDLLENKLQVE 483
 QY 371 -----LKKIQIDIRSLSEEEK-----ELMRIQVDSNPLSEKKEFKL-----KLK 413
 Db 484 ASDWYLELTKHQEPIICKKQOEEMLKQITTELEKEMNIRDELESYRKFEIQQGDEVCK 543
 QY 414 LDIO-----PYDINORLQDPTGLDPSINIDYRKQYKRIQIONIDALHQ-----SIG 461
 Db 544 LDKSEBNARSIEYEVCLK -EKQMILENKCNL -KKQIENSKNIEE-LHQENKALKK 599
 QY 462 SLYLAK -LYLENNNINNLATGADLVSDTDTKIN-KGIFNEKKNKFY-SISSNMI 518
 Db 600 SSAEKQOLMAAE -IKVKLELELA-----STKQFPEMINNYOKLEIKKISEKIL 650
 QY 519 VDINERPLADNERIKWRIQISPTPRAGY-----LENGKILQINIGLEIKDQIIOK 570
 Db 651 GEVEKAKATVDEAVK--LQKEIDLRCQHKIAEVALMEKKHODKYIVERSESELGLKN 708
 QY 571 SEKEYIRIDAKVVRKSIDTKIOE-----AQLINQEMKKAIGL -PKYKTLITFVNHNRY 624
 Db 709 REOE--QSSAKVALETETSLNRIEVLKQLQVEKEEKEKLMQEDENTAILTDKDKKI 766
 QY 625 ASNIYESAYLLNBMK---NNIOSDLKRYNTLVNG-----RPFYDTITPLNIAE 674
 Db 767 QASLIESE -ATMKKDSKTTPSQNISIRLSSMSDGSKSKDNBSIRASAKISLTVYTK 824
 QY 675 QYTHODELYEQVHSGKGLVPESRSILHGPSK-----GVELRNDSE-----GFIEHFG 722
 Db 825 EYT-----VKPTPKKSIYQENBKTLPGSGNKKRKATVPFEDVNDOSSETTDLSLVSE-- 877
 QY 723 HAVDYAGYLLDKQSD--LYTNSKRTIDIFKEGSGNLSTYG 762
 Db 878 --EDISNRIYNNPTPSHLLVKTPKQTPLSLSTPASFTKFG 916

RESULT 12
 S49461
 synaptotemal complex protein 1 - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-Nov-1999
 C:Accession: S49461; S59599
 R:Julien, S.; Luc, M.; Francois, C.
 submitted to the EMBL Data Library, October 1994
 A:Description: Cloning and sequencing of the murine SCP1 cDNA.
 A:Accession number: S49461
 A:Reference: S49461
 A:Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 1-993 <JUL>
 A:Cross-references: EMBL:Z88118; NID:g1360015; PIDD:CA86262.1; PID:g55860303
 R:Saage, J.; Martin, L.; Cuzin, F.; Rassoulzadegan, M.

Biochim. Biophys. Acta 1263, 258-260, 1995

A:Title: cDNA sequence of the murine synaptonemal complex protein 1 (SCP1).

A:Reference number: S59599; MUID:96004899

A:Accession: S59599

A:Status: Preliminary

A:Molecule type: mRNA

A:Residues: 1-993 <SAG>

A:Cross-references: EMBL:Z38118; NID:q1360015; PIDN:CA86262.1; PID:q558603

Query Match: 5.7%; Score 228.5; DB 2; Length 993;

Best Local Similarity 22.0%; Pred. No. 0.038;

Matches 202; Conservative 144; Mismatches 336; Indels 237; Gaps 49.

44 VEEKENKNDNRKDEERKKTQEEHLKELMKHIVIEVGEAAV----- 88

124 LKQKKNKIQDENKKTIEAQRA-----IQELQFENEKYSLKLEELIQENDLIKENNATIM 179

89 ---KEAAELKLEKVPDVLNEMYKAIGKIYI-VDDITKHI---SLEALSDDK--- 135

180 CNLLETGARSAAEK---TNKYEREETROYVDLNSINIEKMLAEELRVQENARLEKH 237

136 KIKIYCKDALNHHY---VYAKGEYDVLIQSSDEVVENTKALNAYVIEGKILSD 191

238 FKLDKDEHKIQHLEETQEVKNKKNQVSELQISKEK---EKMKDLTFLE---ESKD 291

192 ILSKINQYQKE-----LDVLTIKNASDSGDGL-----LFTNOLKEHPTDFSVF 238

292 ---KANQLEFKTKLQDENLKESEKKDHLTSELEDIKMSQMSYQKLEEDLQATKT 348

239 LEQNSNEVQ-----VPAKFAFYIE-----PQHNDVQLVLP 271

349 ISQ-LTEVKEAQMEELNKAKTHTSFVVTVELKATCTLEELTLTEQORLEKKNDDQKLTV 407

272 E-----AFNYWDKE-NEOEIMLSLEELK-----DQRMLSREKKEKIKQNHQSDLS 319

408 ELQKKSNELEMTKTKFNKKEV---ELEELKNILAEQKLLDEKKQYKEKLAELQEKKEQELT 465

320 ---EEGRGLKKLQIPIEPKKDDITHLSQ-----EKELKRIQIDS--DFLSTEEK 368

466 FLLETREREVHDLQAVYVTKTSEQHYLKQVEEMKTELEKELKMTLTAASCDMLLENK 525

369 EF-----LKKLQIDIRDSLSEEEKILNRIQVDSNPLSEKE-----KE 407

526 KTVQASDMALELKKHQEDITNCKQOER-LKQLE-----NLEKEMLRDELESYKRE 579

408 FLK-----KLKIDQPYDINQ-----RLQDTGGLIDSPSINLDVRYKQYKRDQNTALL 456

580 FLOQDEYVCKLKDKESEENARSIECEVLKKEKQMKLIESKCNML--KKQVENSKNSE-L 636

457 HOSIGSTLYNK-----IYKENNNINPLATIGADLVDSDNKIKNGTNEFKKNK 509

637 HOE-NKTLKKSSAEIKQNLNAYE-TRVSKLELEL-----ESTKQR---FEEMTNNQ 683

510 ---YSISNMYIDINERPALDENRLEKWRIQLSPTDFRAGY-----LENGKLIILOR 556

684 KEIEKKKISIEGKLGEVEKAKATVDQAVK--LQKEIDLRCQKRIAEWALMEKHKHQYQK 741

557 NIGLEIKDQVITIKQSEKEYIRIDAVVPRKSIDTKIQE-----AQLTNQENKAKGLPK 611

742 IVEEEDSEELGYKKNQOE--OSSAKIALETTELSTNIRNELVLSKLQLEIEKEKEKLMKAK 799

612 -YTKLITFNVHNRVANSIYESAYLILNEK---NNIQSLDKKYNVYLDG-----NG 660

800 ENTATLKKDKKQKIDQASLLESE--ATSMKPKFSKPTTPSGNISRSLSSMDQSKSDRNL 857

661 RFVFTDITLPLNIAEQYTHODEIYEQVNH--GLVDESRSILLHGSPKGYELANDSE--- 715

858 RASAKSILPTVTYKTEYVATPKTKSIQGENKYIFPGSGNKKRKRTAFEDVDVSDSETD 917

716 ---GTFHEGHADVAGVILDKNSD---LVYNSK-----KFDIFPEEGN 757

918 LLSLVSE-----EDVSNRLYDNNPPDSHLVLTPTQTPLSLSTPASFMFGSLKKMREDR 972

OY 758 LTVSGRTNEAFPAFAERL 776
 Db 973 WTIIAKIDRRKRRLKEAEKL 991

RESULT 13

RESA-3763
 RESA-3763 antigen PFB0915W - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
 C:Accession: B71603
 R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 ; Perera, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;
 Science 282, 1126-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; M01D:99021743
 A:Accession: B71603
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1558 <GAR>
 A:Cross-references: GB:AE001424; GB:AE001362; MID:g3845307; PIDN:AAC71972.1; PID:g3845307
 A:Experimental source: clone 3D7
 C:Genetics:
 A:Gene: PFB0915W

Query Match 5.6%; Score 224.5; DB 2; Length 1558;
 Best Local Similarity 18.5%; Pred. No. 0.096;
 Matches 156; Conservative 123; Mismatches 260; Indels 305; Gaps 29;

OY 44 VKEKKNKDEKRRKDEERNKTOEHLKEIMKHIVKIEVGKEEAVKKKAELKEV----- 99
 Db 659 IEKKLELHNHNVSALENTQSEKKEVIDV-----EEYKEVATTLIEYEQAE 710
 OY 100 ---PSDVLKMYKAIGKITYVDGDTKHT-----SLEALSE 132
 Db 711 EESASTTEIFENILFENNAVESNENVAENLEKNETVFNTLDKVEETVEISGESLENEM 770
 OY 133 DKKKIDYIGKDLHHEHYVAK-EGYEPVLYQSS-----DYVENEKAL 178
 Db 771 DKAFSEIFDNVKGIGQENLITGMFRSIEISIVIOSEKVDLENVSSILDNTEENKEGL 830
 OY 179 -----NYYEIGKILSRD-----ILSKINQYOKFLDVLNTIK 211
 Db 831 LNKLENISSTEGQVETVEHVEBNVYVDVPRMKQOFLGILNAGLKEKMFNLEBYEK 890
 OY 212 NASDSGGDLFTNQLKEHPTDSVEFLQNSNEVOEFAKAFAYIIEPQHRDYLQYAP 271
 Db 891 SES-----DVITVEEIKDEP-----VQKEVEKETVSIIEEMENIVDLLEE 931
 OY 272 EAFNYMDKF-----NEQELNSLEELKQRMISREKEMKIK 308
 Db 932 EKEDLTKMIDAVEESIEISSDSKETESTIKDEKDVSLVEEVQDMDSEVEKYLELK 991
 OY 309 QHYQHW-----SDLSSEGRGL-----LKKIQPIE-----PKD 338
 Db 992 NMEELMKDAVEINDITTSKLIETQELNEVEADLIKDEMLTELKALSDSKELIDAK 1051
 OY 339 DIHLSQOEKEL-----LKRQIDIS-----SDFLSTEE-----KEF----- 370
 Db 1052 DTLEKYLEEHDITTLTLEVEVLEKQVEDEKIEKVSCLKDEEDILKEVKEIKELSESTLE 1111
 OY 371 -----LKKLQIDIDSLSEKELLNRIQVDSNPLSEKEKEFLKAKLQIDQPIDINQRLQ 426
 Db 1112 DYKELTTEIDLEEKKEIKDEHFEK-----EEKEEIKLEADL-----LK 1154
 OY 427 DTGGLIDSPINLDRKQYKQRIQNI-----DAL-----LHQSIGSTLYNKIY 469
 Db 1155 EVSSLEVEEKKLEVEYHELKEVEYHIIISGAHIIKGLDEEDLVEVDLKGSIILMLKGDME 1214
 OY 470 L--YENNNINNLATLQADLVSTDMTKINRGIFNEKKKFKKXSSISSNVYIVDINERPAL 527
 Db 470 L--YENNNINNLATLQADLVSTDMTKINRGIFNEKKKFKKXSSISSNVYIVDINERPAL 527

Db 1215 LQDMKESLDEVTAKLGERVESLKDVLSSALGMDQEMKTRKKA-----QRPKL 1263
 OY 528 DNERLKRIQDSPDTRAGVLENGKLIQRIKIGLEIYD----- 564
 Db 1264 EEVLLKEEVEEKP-----KKTTRKVRFDIKKEPEKDEIVEEMKDEDIDEDIEE 1314
 OY 565 -----VQIIQSEKEVYIRIDAKVPSKIDTK 591
 Db 1315 DVEEDIEEDKVEDIDEDIDEDIDEDIGEDKDEYIDILVQKEKREKYEK-----KKKELEK 1371
 OY 592 IQEQLQINQEWKNAKGLPRTYKLTFFNVHNRVASNTVESAYLILNEKNNISDLIKKY 651
 Db 1372 VEEGVSLKRRHVEWM---RYVQKIDKEV-DKEVSKALESKNDVTYVLKON--QDFFSKV 1425
 OY 652 TNYL 655
 Db 1426 KNFV 1429

RESULT 14

T18427
 hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T18427
 R:Lawson, D.; Bowman, S.; Barrell, B.
 Submitted to the EMBL Data Library, August 1997
 A:Reference number: Z18935
 A:Accession: T18427
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3724 <LAN>
 A:Cross-references: EMBL:298547; MID:e1325376; PID:e1325379; PIDN:CAB11104.1
 A:Introns: 307/1; 1545/2
 A:Note: C0335c

Query Match 5.6%; Score 221.5; DB 2; Length 3724;
 Best Local Similarity 19.5%; Pred. No. 0.38;
 Matches 170; Conservative 174; Mismatches 321; Indels 207; Gaps 39;

OY 45 KEKKNKDEKRRKDEERNKTOEHLKEIMKHIVKIE-----VKGE 84
 Db 497 KELOERDEKKNKDKITYNNNEQTDOLNINIKIESINNNDNNNNINNKKEKRTRE 556
 OY 85 EAVKKEAEKLEKVP---DVLKMYA-----IGKITYYDGDITK- 123
 Db 557 HILNKESISRHIGSPSRDKKIKLYTNKNESTFELKKELEIITNNKNVVEEDIGS 616
 OY 124 -----HISLEALSED-----KKKIKDIYGDALLHHEHYVAKEGEVPVLYQSSDY 170
 Db 617 NEDEYTHVLKENLKEBANENNDKKNKNTYEILSKVYLEKRTLEBLKRGKNT 676
 OY 171 VENTKALNVYVEIGKILSRDIL-----SKINQ-----PYQFLDV---LNTIKNASDS 217
 Db 677 FKDKD-----YNSIGVYIIEIQINEENKINDIODGINSQKIIOSSSRNDNFENIKDIS 732
 OY 218 GQDLFTNQLKEHPTDSVEFLQNSNEVOEFAKAFAYIIEPQHRDYLQYAFNMY 277
 Db 733 LNDLEKEKRRKKSQHFIDMLYKADKNEISENIKIDNNINNIYDESINNIYDESINNI 792
 OY 278 DKFNEQELN-----SLELKQRMISREKEMKIKQHYQHNSDSISEGRGLKQPIE 334
 Db 793 --YDESINNIYDESINNIYDENINNIYD--ENINNIYDENINNIYDGIKIKIDDNIL-LE 847
 OY 335 PKR-----DIIY-----HSLSQEKELKRIQIDSS-----FLSTEEKEFLKLLQ 375
 Db 848 NKNIKTINDIYQVEENNESLEKNEMLMSLKNDINNIYMKKEVNDIFINIKIKESLAKID 907
 OY 376 IDIRDSLSEKELLNRIQV-----SSNPLS-EKEKEFLKLLQIDQPIYDIN 422
 Db 908 KNIKDNNNDDEYIMDNFYENDFIINHKMEITNKELDPLEINTQNEFFTEINDIKKKKYTND 967

QY 423 ORLODTGLIDSPS--INLDVAKOYKRDIONDALLHOSIGSTLYNKTYL----- 471
Db 968 HFFENADKMEYEMNKILKDKMKNEOEFPKTD-----EFGSLQSHKIKKYKGEKHDK 1023
QY 472 ---ENMNI-----NLTATGADLVDTNTKINRGIFNEFKKFNFYSSNMYIDNER 524
Db 1024 NNEEKKNLYDENOVYVLYSD-----HKTEODIOD-----IHSIQTN--YDENNI 1068
QY 525 PALDNERLKWRIOQSPD--TRAGYLENGKLLQNRNIGLEIKDVQIIOSEKEYIRIDAKV 582
Db 1069 EQINENSKKGVRIKSTDMENKNDMENKNDMEKKNDEKKNDEKKNDEKKNDEKKNDEKKN 1127
QY 583 VPKKIDPKIKQEAQINIQEWKKA-----LGLPKTKILITNV--HNKVASNI 628
Db 1128 DMEKKNDEKKNDEKKNDEKKNDEKKNDEKKNDEKKNDEKKNDEKKNDEKKNDEKKN 1187
QY 629 VESAVLLINEMKNN--IOSDLIK--VTNVLVDGNGRFV-----TDITLPIAE-QYTH 678
Db 1188 IENNNIENNNIENNNIENNNIENNNIENNNIENNNIENNNIENNNIENNNIENNNIENNN 1247
QY 679 QDEIYEQVHSGLYVPESRSILLHGPSKGVLELRD-----SEGFIHEFGHAVDDYAGY 731
Db 1248 ONNIGNKVH-----VKDKKNDLIN-----NVDIINDVLKSDKFEKMINSKETINIKDFK-- 1296
QY 732 LLDKNQSDLVTN-----SKKFIID 750
Db 1297 CVDKN-NDICLNLNELNKNOSTNSNTTEKCID 1327

RESULT 15
670163
hyothetical protein BB0512 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: G70163
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997.
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943
A:Accession: G70163
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2166 <KLE>
A:Cross-references: GB:AE001153; GB:AE000783; NID:q2688419; PIDN:AC66876.1; PID:q268842
A:Experimental source: strain B31

Query Match 5.5%; Score 220; DB 2; Length 2166;
Best Local Similarity 19.8%; Pred. No. 0.22;
Matches 172; Conservative 148; Mismatches 271; Indels 276; Gaps 41;

Db 1325 -----ERCNEGOLNLENKIDNKIKAIIDNALISQY 1353
QY 302 EKME-KIKOHQWMSD-----SLSEGRGLLKKIPIPKKDDIIT--SLSQEKE 350
Db 1354 DGLKRYADWDEFSERLNSYATLSEEFKSSNKRMEFELESQNLKNLSDDLNVKED 1413
QY 351 LKRIQIDS-----SPLSTEEKEFLKIQI--DIRDSLSE---BEKELNRIQVDSNP 400
Db 1414 VI-RLKEESYHNWSSHLLEEDFEFKDLIRGEELKYSLENFASYNKIONLEYDLSKN 1472
QY 401 LSEKEFEFLKKLIDIOFYDINORLODTGGLIDSPSINLDVAKOY---KRDIONDALLH 457
Db 1473 L-ENKTELIOSRDIE-----QKKKD-----DKENFYLDFTKESSKKKQDSIALME 1521
QY 458 QISGSTLYNKIYLYENMN--INLTATGADLVDT-----STDNTKIN-----RG 499
Db 1522 TITGKVDPEFVFNKQSIIDSMFLNIKDVKQWQEKSYSTIEKRIMLAELGKISFEND 1581
QY 500 IFN-----EKKKNKY---SISSNMYVDINERBALNRLKWRIOQSPDTR--AGYLEN 549
Db 1582 IFNWKIGLESFRDGEIIEAEIIFSN-----IQNEAKKIEQSVHLDKFNIGESLN 1630
QY 550 GLKI-LQNRNIGLEIK--DVQIIOSEKEYIRIDAKVVPKSK-IDTKIQEAQINIQEWNK 605
Db 1631 LKVIDLEKFFVDFKLEKIDKYNKKTEDILIQAEVFLQOKDLEKIFELMOKLEHETT 1690
QY 606 -ALGLPKYTK-----LITFNH-----NRYASNI---VESAVYLLN 637
Db 1691 LSSNLDKVRREWVDYISSDKESFEQIILINKINISEFSEKISLYRNNTIETSIENEY--N 1747
QY 638 EMKNNIOSDLIKKVVNVLVDGNGRFVFTDITLPIAEQYTHQDEIYEQVHSGLYVPESR 697
Db 1748 SFSKSKIRDL-----GVELRNDSEGFIEFGHAVDDYAGYLLDKNQSDLV 1765
QY 698 SILLHPSK-----GVELRNDSEGFIEFGHAVDDYAGYLLDKNQSDLV 1765
Db 1766 KSLKSTSEIETIKSGLOEQIDKFEVEFKKNNKELLKEVDNNIILESKILN---CDVOF 1822
QY 743 NSKRFIDFKKEGSLUTSYGRNEAEF 769
Db 1823 N-KFISEIKD--NLVEYKSDLRAEF 1844

Search completed: December 2, 2001, 13:51:02
Job time: 297 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:51:33 ; Search time 44.8 Seconds
(without alignments)
636.724 Million cell updates/sec

Title: US-09-747-521-2_COPY_1_778
Perfect score: 3987
Sequence: 1 MNKEFKIVISMCLVTAL.....TSYGRNNEFFAEFRIMH 778

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3987	100.0	809	1	LEF_BACAN
2	501	12.6	800	1	CYAA_BACAN
3	232	5.8	2869	1	RBP1_PLAVB
4	229.5	5.8	997	1	SCP1_RAT
5	228.5	5.7	993	1	SCP1_MOUSE
6	223	5.6	794	1	HMR_MOUSE
7	219.5	5.5	2663	1	CENE_HUMAN
8	218.5	5.5	1075	1	Y124_METJA
9	215	5.4	2245	1	MG51_DICTDI
10	209.5	5.3	1957	1	YD86_SCHPO
11	209	5.2	1679	1	Y109_YEAST
12	208.5	5.2	691	1	Y104_YEAST
13	208	5.2	1230	1	SMG3_YEAST
14	205.5	5.2	756	1	Y348_MYCGE
15	205.5	5.2	1875	1	MLP1_YEAST
16	202	5.1	2710	1	TOXA_CLODI
17	202	5.1	2748	1	NUM1_YEAST
18	201.5	5.1	3911	1	AKA9_HUMAN
19	201	5.0	1790	1	USO1_YEAST
20	200	5.0	845	1	SCP1_MESAU
21	199.5	5.0	1251	1	RBP2_PLAVB
22	199.5	5.0	2116	1	MY62_DICTDI
23	199	5.0	1312	1	RA50_YEAST
24	197.5	5.0	1427	1	REST_HUMAN
25	197.5	5.0	1928	1	MY51_YEAST
26	197	4.9	1940	1	MYH3_RAT
27	196	4.9	976	1	SCP1_HUMAN
28	196	4.9	1939	1	MYH6_MESAU
29	195.5	4.9	1805	1	HMR_MYCGE
30	194.5	4.9	3210	1	CENE_HUMAN
31	194	4.9	1727	1	ALM1_SCHPO
32	192	4.8	2230	1	GOG4_HUMAN
33	191.5	4.8	944	1	NDF1_YEAST

34	191.5	4.8	1164	1	BAG_STRAG	P27951	streptococc
35	190	4.8	1818	1	HMW2_MYCPN	P75471	mycoplasma
36	190	4.8	1938	1	MYH6_RAT	P02563	rattus norv
37	189.5	4.8	1935	1	MY58_CYPBA	P09039	cyprinus car
38	189.5	4.8	1978	1	MYH8_CHICK	P10587	gallus gall
39	189	4.7	724	1	HMR_HUMAN	P07530	homo sapien
40	187.5	4.7	681	1	RPSD_HELPJ	P09293	helicobacte
41	187.5	4.7	1935	1	MYH7_PIG	P79293	sus. scrofa
42	187.5	4.7	1959	1	MYH9_CHICK	P14105	gallus gall
43	186	4.7	1938	1	MYH6_MOUSE	P02566	mus musculu
44	184.5	4.6	1966	1	MY58_CAEEL	P02566	caenorhabdi
45	184	4.6	1939	1	MYH6_HUMAN	P13533	homo sapien

ALIGNMENTS

RESULT 1
LEF_BACAN
ID LEF_BACAN STANDARD; PRT; 809 AA.
AC P15917;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE LETHAL FACTOR PRECURSOR (EC 3.4.24.-) (Lef).
GN LEF.
OS Bacillus anthracis.
OG Plasmid pXOI.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 34-49.
RX MEDLINE=90034185; PubMed=2509294;
RA Bragg T.S., Robertson D.L.;
RT "Nucleotide sequence and analysis of the lethal factor gene (lef)
from Bacillus anthracis.";
RL Gene 81:45-54(1989).
RN [2]
RP SEQUENCE FROM N.A.
RA Lowe J.;
RL Submitted (APR-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RX MEDLINE=95154669; PubMed=7851740;
RA Kochl S.K., Schlavo G., Mock M., Montecucco C.;
RT "Zinc content of the Bacillus anthracis lethal factor.";
RL FEMS Microbiol. Lett. 124:343-348(1994).
CC -1- FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,
DEATH. LEF IS THOUGHT TO BE A LETHAL FACTOR THAT, WHEN ASSOCIATED
WITH PA, CAUSES DEATH. LEF IS NOT TOXIC BY ITSELF. PA IS THOUGHT TO
BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC CELLS, THEREBY
FACILITATING THE INTERNALIZATION OF LEF OR EF.
CC -1- SUBUNIT: SECRETED ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT
PROTEINS. A PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN
EDEMA FACTOR (EF). NONE OF THESE IS TOXIC BY ITSELF.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: THE PA-BINDING REGION IS FOUND IN BOTH B. ANTHRACIS EF
AND LEF.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M34 (ZINC
METALLOPROTEASES).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29081; AAA79216.1; -
DR EMBL; M30210; AAA22569.1; -

DR PIR: JQ0032, JQ0032.
 DR MEROPS: M34..001; Anthrax toxin.
 DR InterPro: IPR003541; Anthrax toxin.
 DR InterPro: IPR000130; Zn.MTpeptidase.
 DR PROSITE: PS00142; ZINC_PROTEASE_1.
 KW Hydrolyase; Metalloprotease; Zinc; Toxin; Signal; Repeat; Plasmid.
 FT SIGNAL 1 33
 FT CHAIN 34 809 LETHAL FACTOR.
 FT DOMAIN 34 293 PA-BINDING REGION (POTENTIAL).
 FT METAL 300 420 REPEATS.
 FT METAL 719 719 ZINC (CATALYTIC) (POTENTIAL).
 FT ACT_SITE 720 720 POTENTIAL.
 FT METAL 723 723 ZINC (CATALYTIC) (POTENTIAL).
 SO SEQUENCE 809 AA: 93786 MW: 8C16B4D727310AB CRC64;

Query Match 100.0%; Score 3987; DB 1; Length 809;
 Best Local Similarity 100.0%; Pred. No. 5.2e-148;
 Matches 778; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKKEFIKIVSMGCLVTAITLSCGPVPIPVQAGAGHDVGMMHKEKKNDEKRNKDEE 60
 1 MNKKEFIKIVSMGCLVTAITLSCGPVPIPVQAGAGHDVGMMHKEKKNDEKRNKDEE 60
 DB 1 MNKKEFIKIVSMGCLVTAITLSCGPVPIPVQAGAGHDVGMMHKEKKNDEKRNKDEE 60

QY 61 RNKQOEHLKELMKHYIVIEVKEGEAVKKEAEKLEVPDVLKMYAIGKITYVDG 120
 61 RNKQOEHLKELMKHYIVIEVKEGEAVKKEAEKLEVPDVLKMYAIGKITYVDG 120
 DB 61 RNKQOEHLKELMKHYIVIEVKEGEAVKKEAEKLEVPDVLKMYAIGKITYVDG 120

QY 121 ITHKISLESEDKKIKIDYIGKDALHEHYVYAKEGEPVLVIOSSDYVENTKALNV 180
 121 ITHKISLESEDKKIKIDYIGKDALHEHYVYAKEGEPVLVIOSSDYVENTKALNV 180
 DB 121 ITHKISLESEDKKIKIDYIGKDALHEHYVYAKEGEPVLVIOSSDYVENTKALNV 180

QY 181 YVEIGKILSRDILSKINPYOKFLDVLNTIKNASDSQDGLFTNQLKHPDTSVEFL 240
 181 YVEIGKILSRDILSKINPYOKFLDVLNTIKNASDSQDGLFTNQLKHPDTSVEFL 240
 DB 181 YVEIGKILSRDILSKINPYOKFLDVLNTIKNASDSQDGLFTNQLKHPDTSVEFL 240

QY 241 QNSNEVOVFKAAPAYIEPQHRVLYLAPBAFVYMKFNEQENLSLEELKQRMISR 300
 241 QNSNEVOVFKAAPAYIEPQHRVLYLAPBAFVYMKFNEQENLSLEELKQRMISR 300
 DB 241 QNSNEVOVFKAAPAYIEPQHRVLYLAPBAFVYMKFNEQENLSLEELKQRMISR 300

QY 301 YEKKEKIKQHQHMSDSISEGRCGLKQIPIEPKDDIHSLSQSEKELKRIQIDSS 360
 301 YEKKEKIKQHQHMSDSISEGRCGLKQIPIEPKDDIHSLSQSEKELKRIQIDSS 360
 DB 301 YEKKEKIKQHQHMSDSISEGRCGLKQIPIEPKDDIHSLSQSEKELKRIQIDSS 360

QY 361 DFLSTEKEFLKQIQRDLSLSEEEKELNRIOVDSNPLSEKEFEFLKQIQRD 420
 361 DFLSTEKEFLKQIQRDLSLSEEEKELNRIOVDSNPLSEKEFEFLKQIQRD 420
 DB 361 DFLSTEKEFLKQIQRDLSLSEEEKELNRIOVDSNPLSEKEFEFLKQIQRD 420

QY 421 INORLODTGGLIDSPINLDVRKQYKRDIONIDALHOSIGSTLYNKLYENMINMLT 480
 421 INORLODTGGLIDSPINLDVRKQYKRDIONIDALHOSIGSTLYNKLYENMINMLT 480
 DB 421 INORLODTGGLIDSPINLDVRKQYKRDIONIDALHOSIGSTLYNKLYENMINMLT 480

QY 481 ATLGADVDSTDNTRKINGIFNEFEKKNKYSSNMYIVDINERPALDNERLKRIQISP 540
 481 ATLGADVDSTDNTRKINGIFNEFEKKNKYSSNMYIVDINERPALDNERLKRIQISP 540
 DB 481 ATLGADVDSTDNTRKINGIFNEFEKKNKYSSNMYIVDINERPALDNERLKRIQISP 540

QY 541 DTRAGYLENGKLLIQRNIGLEIKDVQIKQSEKEYIRIDAVVPKSKIDTRKIQEAQLIN 600
 541 DTRAGYLENGKLLIQRNIGLEIKDVQIKQSEKEYIRIDAVVPKSKIDTRKIQEAQLIN 600
 DB 541 DTRAGYLENGKLLIQRNIGLEIKDVQIKQSEKEYIRIDAVVPKSKIDTRKIQEAQLIN 600

QY 601 QEMKKAIGLSPYTKLIFPNVNNRYSANVESAYILLNEMKNKNIOSDLKKKTYNLVDNG 660
 601 QEMKKAIGLSPYTKLIFPNVNNRYSANVESAYILLNEMKNKNIOSDLKKKTYNLVDNG 660
 DB 601 QEMKKAIGLSPYTKLIFPNVNNRYSANVESAYILLNEMKNKNIOSDLKKKTYNLVDNG 660

QY 661 RFVFTDITLPIAIOYTHODEIEQVSHKGLVPEPSRIILHGPCKGYELNDSGCFIHE 720
 661 RFVFTDITLPIAIOYTHODEIEQVSHKGLVPEPSRIILHGPCKGYELNDSGCFIHE 720
 DB 661 RFVFTDITLPIAIOYTHODEIEQVSHKGLVPEPSRIILHGPCKGYELNDSGCFIHE 720

QY 721 FGHAVDVAGYLLDKNOSDLVNSKFFIDIFKEGNSLTSYGRNEAFFAEAPFLMH 778
 721 FGHAVDVAGYLLDKNOSDLVNSKFFIDIFKEGNSLTSYGRNEAFFAEAPFLMH 778
 DB 721 FGHAVDVAGYLLDKNOSDLVNSKFFIDIFKEGNSLTSYGRNEAFFAEAPFLMH 778

RESULT 2
 ID CYAA_BACAN STANDARD; PRT; 800 AA.
 AC P40136;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CALMODULIN-SENSITIVE ADENYLATE CYCLASE PRECURSOR (EC 4.6.1.1) (ATP
 DE PYROPHOSPHATE-LYASE) (ADENYLYL CYCLASE) (EDEMA FACTOR) (EF).
 GN CYA.
 OS Bacillus anthracis.
 OC Plasmid pXOI.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1392;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89138004; PubMed=2906312.
 RA Escuyer V., Duflot E., Sezer O., Danchin A., Mock M.;
 RT "Structural homology between virulence-associated bacterial adenylate
 RT cyclases.";
 RL Gene 71:293-298(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89211974; PubMed=3149607;
 RA Robertson D.L., Tippetts M.T., Leppia S.H.;
 RT "Nucleotide sequence of the Bacillus anthracis edema factor gene
 RL (cya): a calmodulin-dependent adenylate cyclase.";
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Escuyer V., Duflot E., Mock M., Danchin A.;
 RT "Nucleotide sequences expressing adenylate cyclase from B.anthraxis,
 RT proteins having the activity of this adenylate cyclase and biological
 RT uses.";
 RL Patent number EP0366550, 02-MAY-1990.
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88198021; PubMed=2834337;
 RA Tippetts M.T., Robertson D.L.;
 RT "Molecular cloning and expression of the Bacillus anthracis edema
 RT factor toxin gene: a calmodulin-dependent adenylate cyclase.";
 RL J. Bacteriol. 170:2263-2266(1988).
 RN [5]
 RP SEQUENCE OF 34-48.
 RX MEDLINE=89211974; PubMed=3149607;
 RA Schmidt J.;
 RT Unpublished results, cited by:
 RL Robertson D.L., Tippetts M.T., Leppia S.H.;
 RN [6]
 RP MEDLINE=91119764; PubMed=8418825;
 RA Danchin A.;
 RT "Phylogeny of adenylyl cyclases.";
 CC Adv. Second messenger Phosphoprotein Res. 27:109-162(1993).
 CC -1- FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,
 CC AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE
 CC DEATH. EF IS A CALMODULIN-DEPENDENT ADENYLYL CYCLASE THAT, WHEN
 CC ASSOCIATED WITH PA, CAUSES EDEMA. EF IS NOT TOXIC BY ITSELF. PA IS
 CC THOUGHT TO BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC CELLS,
 CC THEREBY FACILITATING THE INTERNALIZATION OF EF OR LF.
 CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.
 CC -1- ENZYME REGULATION: ACTIVATED BY HOST CALMODULIN.
 CC -1- SUBUNIT: ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT PROTEINS, A
 CC PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN EDEMA FACTOR
 CC (EF). NONE OF THESE IS TOXIC BY ITSELF.
 CC -1- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-2 FAMILY.
 CC AND LF.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M23179; AAA22374.1; -
DR EMBL: M24074; AAA79215.1; -
DR EMBL: A07289; CA00652.1; ALT-SEQ.
DR Interpro: IPR003541; Anthrax_toxinA.
KW Lyase; CAMP synthetase; Toxin; ATP-binding; Signal; Plasmid.
FT SIGNAL 1 33
FT CHAIN 34 800 CALMODULIN-SENSITIVE ADENYLATE CYCLASE.
FT DOMAIN 34 288 PA-BINDING REGION (POTENTIAL).
FT NP_BIND 289 680 CATALYTIC.
FT NP_BIND 347 354 ATP (POTENTIAL).
FT CONFLICT 350 350 V -> E (IN REF. 2).
FT CONFLICT 510 510 Q -> T (IN REF. 2).
FT CONFLICT 512 513 EW -> RM (IN REF. 2).
FT CONFLICT 760 760 V -> L (IN REF. 3).
SQ SEQUENCE 800 AA: 92477 MW: 94758485DF4C5A6 CRC64;

Query Match 12.6%; Score 501; DB 1; Length 800;
Best Local Similarity 24.6%; Pred. No. 5.6e-13;
Matches 202; Conservative 145; Mismatches 281; Indels 194; Gaps 36;

OY 4 KKEFT---KVISMCLVATITLSCVPFIPVLOGAGHGDCVGMHVKKEKNKDKRDE 59
DB 3 RKFEPNKRSTIISFVYLRAISSQAIEVNA-----MNEHTESDIKRNHKT 49
OY 60 ENRKOEHLEKMKHIVIEVKGGEAVKKEAEKLEVPDSVLYMKAIGIKIYVG 119
DB 50 ENRKEKEKFKOSINLVTEFTNELDKIQOTDLKKIKPVOLLEYSELGEIYFTDI 109
OY 120 DITKHISLEALSEDKKKIKDIYKDALLHEHYVYAKEGEYEVLYIOSEEDYVNTKALN 179
DB 110 DLVEHKELODSEEEKSNMNSNGEKVPASRFYEKKREPKILI-NIKDYAINSSQSK 168
OY 180 VYEYEGIKLSRLSKINPYOKFLD-VLNTIKNNS-DSGODLLFTNOLAK---HPT 222
DB 169 VYEYEGIKLSRLSKINPYOKFLD-VLNTIKNNS-DSGODLLFTNOLAK---HPT 222
OY 223 DRSVEFLQNSNEVOEFAKAFAYIEPOHRYDLYOLYAPAEVYMDKPNQNEINLSLEL 292
DB 224 SIDINFIKENLPEFOHAFSLAFSYFAPDRHYLEYIAPDMFYMKL----- 271
OY 293 KDQRLMRYEK-WEKIKOHYOHMSDSLSEEGRLKKLQIPIEPKDDIINLSOEEKE 350
DB 272 -----EKGEFEKI-----SESLKKEG-----VEKDRIDVL---KGEKA 301
OY 351 LAKRIOIDSSDPLSTBEKEFLKKLOIDIRDSLSEEEKELNKRQVUNSSNPLSEKEFEK 410
DB 302 L-----KASGVLPEHAAPFKI-----ARELNTYLIFFRPVKNLATN---L 339
OY 411 KKLIDPIODINORLODTGGLIDSPSINDVRYKQY-----KRDIONIDALL-HO-STG 461
DB 340 KSGVATKGLNHNHCKSSDKMPVACIYIPFDOLSKKHGOOLAVEKGNLENKKSITIEHGEIG 399
OY 462 SYLYNRIYL-YENMNINNLATLGLADLVSTDNKTINRGIFNEKFNKFK-----YSISNN 515
DB 400 -----KIPKLHLRIEELK-----ENGILLKKEKIDNKKYVILLESN 438
OY 516 YHIVIDINEPALDNERLAKRIQLSPTFRAGYLE-NGKLLQNRNIGLEIDQV-TIQOSEK 573
DB 439 NOYVEE--RISDENNEVOYK-----TKEGKIIVLEGEKFMWRNIEVMKAVEGVLPLTA 490
OY 574 EYIRIDAKVVPKSIDTKIOEAQLNIOENKRALGLPKYTK-----LITFENVNRYA 625
DB 491 DY---DLFALAPSLTEIKKOIPO---KENDKVVNTPNLSLEKOKGYTNLLIKIGIERK-- 541
OY 626 SNVESAVYLILNMKNINIOSDLIKKVTNYLVYDNGRPFVFTDITLPIAEOYTHOD--EIV 683
DB 626 SNVESAVYLILNMKNINIOSDLIKKVTNYLVYDNGRPFVFTDITLPIAEOYTHOD--EIV 683

DB 542 ---PDSNFKGTLNMWQKMDRL-----NEAVKTYGTGVDVNVNHTGEQDNDEFP 587
OY 684 EYHSGKGLVYPERSTILLG---PSKGVELRNDSEGFIEFGVANDYA---GYL----- 732
DB 588 EKDNFEIINDEEFILTKNMETGRFIEKNINIGKYLVEYFNRSYNKAIAPGNKAYIEMTD 647
OY 733 -LDKNOSDLVYNSKRFIDIKKEGSNLTSGRTNEAFEEFAEA 773
DB 648 PITKAKINITPTSAEFT-----KNLSSIRKSSNWGVYKDS 682

RESULT 3
RBP1_PLVAB
ID RBP1_PLVAB STANDARD: PRT: 2869 AA.
AC 000798;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
GN RBP1
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RX MEDLINE=92315338; PubMed=1617731;
RA Galluski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites.";
RL Cell 69:1213-1226(1992).
CC -I- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -I- SUBUNIT: HOMODIMER (POTENTIAL).
CC -I- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M88097; AAA29743.1; -
DR HSSP: P36956; IAM9.
KW Malaria; Receptor; Signal; Transmembrane.
FT SIGNAL 1 17
FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
FT DOMAIN 18 2807 EXTRACELLULAR.
FT TRANSMEM 2808 2826 POTENTIAL.
FT DOMAIN 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2869 AA: 330213 MW: 89DBE442205BECF CRC64;

Query Match 5.8%; Score 232; DB 1; Length 2869;
Best Local Similarity 18.8%; Pred. No. 0.065;
Matches 190; Conservative 159; Mismatches 315; Indels 346; Gaps 46;
OY 39 DVGAMHVKKEKNKDKNR-----KDEERNKTOGEHLEKMKHIVXI----- 79
DB 746 DINALIEVEKFTVENKESTLEMLKDBE---WEKLODKKEFFAKLNPVSDKLDIVYT 801
OY 80 ---EYKGEAVKKEAEKLEKLVPSDVLBYKAIGKI-----YIVGD-ITKAI 125
DB 802 KMSAEVYTNAGIKKEIAQOFENVHKKMKESDAFSPKFPALQNSMQVYNOGEDDAIEKKH 861
OY 126 SLEALSDEK---KKIKDIYKDALLHEHYVYAKEGY---EPVLYIOSEEDYVNTK 176
DB 862 QNRSEKEEYFKNMSVEEDLSRETEEOEYTKRKNNSRRKSGISAETITMREVINKIES 921
OY 177 ALNVEYEEI-----GKILSRDLISKINQYQKFLDLVNLTKNA--- 213

Db 475 REKEIDHLEQVAVTQTESEHLYKQVEEMKTELEKLNIELTANSDMLLENKRLVQE 534
QY 371 -----LKKLOIDIDISLSEEEK-----ELNRIOVDSNPISSEKEPELK-----KLK 413
Db 535 ASDMVELLEKKNHODIINCKKOEERMLKOJETTELEKEENLNLEDELESVKEPFIQOGDEVKCK 594
QY 414 LDIO-----PYDINORLODTGGLDPSINLDVRKQYKRDIONDALLHQ-----SIG 461
Db 595 LQDSSENANSIEYVLK--EKOKILLENKCNLN--KKOLENSKNLEE-LHQENKALKKK 650
QY 462 STLYNK-IYLYENMININLTATGADLVSTDTKIN-RGIFNEPKNFKY-SISSNYMI 518
Db 651 SSANENQOLNAYE-IKYNKLELELA-----STKOFEEMINNYOKEIKKISEKKL 701
QY 519 VDIERNPALDNEPLKWRIOQLSPTRAGY-----LENGKLLORNIIGLEIKVOQIIQ 570
Db 702 GEVEKAKAVYDEAVK--LQKEIDLRQCHKIAEVVALMEKHKHQYDKIVEERDSELGAKN 759
QY 571 SEKEYIRIDAKVYPKSKIDPKIOE-----AOLNIOENMNALGL-PKYTLITFENYHNR 624
Db 760 REOE--QSSAKVALTELSTIRNELVSLKKOLEVEKEEKLMEODENTALTDKDKDKTI 817
QY 625 ASNIVESAYLLLENMK--NNIOSDLIKVTNYLVONG-----RFVFTDITLPNIAE 674
Db 818 QASLESPE--ATSMKFDSKTPSQNISRSLSSMDCKSKONRDSLRASAKSLSTVTVK 875
QY 675 QYTHODEIYEQVHSGKLYPESSILHSGK-----GYELKNDSE-----GFIHEFG 722
Db 876 EYI-----VKTPTKRSIYOREENYLPPTGSGSKRKVTFEEDVSSSETDLSLVE-- 928
QY 723 HAVDYAGYLLDKRQSD--LVTSKKKFFIDFKREGSNLTYSYG 762
Db 929 ---EDISNRINYNNTPDSHLVTPKOTPLSLSTPASFTKFG 967

RESULT 5
SCPI_MOUSE STANDARD: PRT: 993 AA.
ID SCPI_MOUSE
AC 062209: 062329: 009205: P70192:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SYNAPTONEMAL COMPLEX PROTEIN 1 (SCP-1 PROTEIN).
GN SCPI OR SCPI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=CBA; TISSUE=Testis;
RX MEDLINE=96004899; PubMed=7548215;
RA Sage J., Martin L., Guzin F., Rassoulzadegan M.;
RT "cDNA sequence of the murine synaptonemal complex protein 1 (SCPI).";
RL Biochim. Biophys. Acta 1263:258-260(1995).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS; TISSUE=Testis;
RA Kerr S.M., Taggart M.H., Lee M., Cooke H.J.;
RL Submitted (Apr-1995) to the EMBL/GenBank/DBJ databases.
RN 13
RP SEQUENCE OF 1-149 FROM N.A.
RC STRAIN=C57BL/6;
RA Sage J., Li Y., Martin L., Mattei M.-G., Guenet J.-L., Liu J.G.,
RL Hoog C., Guzin F., Rassoulzadegan M.;
RL Submitted (Jan-1997) to the EMBL/GenBank/DBJ databases.
RN 14
RP SEQUENCE OF 95-787 FROM N.A.
RC STRAIN=ICR; TISSUE=Testis;
RA Tsuchida T., Nishina Y., Nozaki M., Uchida K., Nishimune Y.;
RL Submitted (Nov-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF

CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS
CC CHROMOSOMES DURING MEIOTIC PROPHASE.
CC SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF
CC -1- SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.
CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS
CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE
CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE
CC SYNAPTONEMAL COMPLEX (BY SIMILARITY).
CC -1- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,
CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL
CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z38118; CA86262.1; -.
DR EMBL: LA1069; AAA64514.1; ALT_INIT.
DR EMBL: U62864; AAC53335.1; -.
DR EMBL: U62860; AAC53335.1; JOINED.
DR EMBL: U62861; AAC53335.1; JOINED.
DR EMBL: U62862; AAC53335.1; JOINED.
DR EMBL: U62863; AAC53335.1; JOINED.
DR EMBL: D88539; BAA13639.1; -.
DR MGD: MGI:105931; Sycp1.
KW Nuclear protein; Meiosis; Cell division; Phosphorylation;
KW DNA-binding; Coiled coil.
FT DOMAIN 12 97 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 104 815 COILED COIL (POTENTIAL).
FT DOMAIN 114 117 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 697 700 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 898 901 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 978 986 ARG/LYS-RICH (BASIC).
FT CONFLICT 527 527 F -> L (IN REF. 2).
SQ SEQUENCE 993 AA; 115962 MW; 1AAFA90D64FAFE6 CRC64;

Query Match 5.7%; Score 228.5; DB 1; Length 993;
Best Local Similarity 22.0%; Pred. No. 0.026;
Matches 202; Conservative 144; Mismatches 336; Indels 237; Gaps 49;

QY 44 VKEKKKNDKRRKDEBRNKTQEBHKEIKMKHYIKYKGEBAVK----- 88
Db 124 LKQKENKLOENRKIKIEQORKA---IQELOFENKEVSLKLEELIQENKDKLEKENNATIHW 179
QY 89 -----KEAAKLELVKVPDVAEMKKAIGKIIYI--VDGDTKHI-----SLALSDDK---- 135
Db 180 CNLKEKTCARSAEK--TNKYEYERETROYVDINSIEMKILAFEEELRYQAEARLEMH 237
QY 136 -KIKIDYKDALIHEHY---VYAKEGEYPAVLQSSDEYVENTEKALNYVEIGKILSRD 191
Db 238 FKLEDEHEKIQHLEEVQKVENKNKENGVSLLQSAK--ENMKKDLTFLE-----ESRD 291
QY 192 ILKKNQYQKF-----LVLTNTIKNASDSGODL-----LFTQKLEHPDSEVF 238
Db 292 ---KANOLEKTKLQDENLKEKSKDHLTSELEDIMSMQSRMSTOKALEEDLOLATKT 348
QY 239 LEQNSNEVOE-----VFAKAFAYYE-----POHRVLOLYAP 271
Db 349 ISO-LTEVKAQOEELKAKATTISFVVTETLAKATCTIEELRTQOEORLEKNEDOLKITV 407
QY 272 E-----AFNYMDKF--NEQINLSLEELK-----DQRMLSRYEWEKIKOHYHMSDLS 319
Db 408 ELQKKSNELEMKTKFNKKEV--ELEELKNILABDQKLLDEKKQVEKLAELQEKBEELT 465
QY 320 ---EGRGLIKKIQIPIEPKKDIIHSLQE-----EKELLKRIQIDSS--DFLSTEEK 368
Db 466 FLLETREREVHDLQEOVAVTQTSQHYLQVEEMKTELEKELKNIETLASCMLLENK 525

QY 369 EF-----LKKLOIDIDBSLSEKELLNRIQVSSNPLSKE-----KE 407
DB 526 KFOVEASPMALCKKHQDDIINCKKOEER-LIKQIE-----NLEEKHRLDELESVKE 579
QY 408 FLK-----KLKLDIOPYDINO-----RLQDTGGLIDSPSINLVRKQYKRDIONIDL 456
DB 580 FIQGGDEVKCKLDSSEENARSLTECEVLKKEKQMKLSEKCNL--KKQYENKSKIEE-L 636
QY 457 HQSIGSTLYNK-----IYENNNINNLPTATGADLVSDTNTKINRGJFNEKKNF 509
DB 637 HQE-NKTLKKSSAEIKOLNAYE-IKVSKLELEL-----ESTKQR---FEEMTNVYQ 683
QY 510 -----YSISSNMYMIDINERPALDNERLKWRIQLSPDRAG-----LNGCLILQIR 556
DB 684 KEIENKKTISECKLGEVKEKAKATYDEAVK--LQKEIDRCOHKIAEWALMEKKHHOYDK 741
QY 557 NIGLEIKDVQIIOSEKEEYIRIDAKVPSKIDFTIOE-----AQLNINOENKALGLPK 611
DB 742 IVEERDSGLGKKNRQE--QSSAKIALETLSNIRNELVSLKQLEIKEKEKELKMAK 799
QY 612 -YTKLITFNHNRVANSNIVESAYLILNEMK--NNIQSDLIKVTNYLVG-----NG 660
DB 800 ENTAIILKKKKKKIQASLESPE--ATSMKFDKSTPPQSNISRLSSMDGSKNDNDL 857
QY 661 RFVFTDILPNIRABQYHODEIYEQVHSK-GIYVPSRSILHGSKEVELRNDSE----- 715
DB 858 SASAKSIPTTVTKETVTKPTKSTIYORENKYIPTGGSNKKRTAFEDVNSDSSETAD 917
QY 716 --GFIFHGHAUVDDYAGYILDKNOSD--LVYNSK-----KFIDIFEKESGN 757
DB 918 LLSLVSE-----EDVSNLKYNNPPDSHLVTKPQTPLSLSTPASFKFQSLKMKRDR 972
QY 758 LTSIGRTNEAEFFAFRL 776
DB 973 WTIKIDRKRLKEAKL 991

RESULT 6
HMMR_MOUSE
ID HMMR_MOUSE STANDARD; PRT; 794 AA.
AC 000547;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYALURONAN MEDIATED MOTILITY RECEPTOR (INTRACELLULAR HYALURONIC ACID BINDING PROTEIN) (RECEPTOR FOR HYALURONAN-MEDIATED MOTILITY).
GN HMMR OR IHABP OR RHAMM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090.
RN [1]
RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
RC TISSUE=Lung;
RX MEDLINE=98264863; PubMed=9601097;
RA Hofmann M., Fieber C., Assmann V., Goettlicher M., Sleeman J., Plug R., Howells N., von Steirn O., Ponta H., Herrlich P.;
RT "Identification of IHABP, a 95 kDa intracellular hyaluronate binding protein";
RT J. Cell Sci. 111:1673-1684(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Zhao Y., Zhang S., Turley E.;
RL submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-183 FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=129/SV;
RX MEDLINE=99107769; PubMed=9889313;
RA Fieber C., Plug R., Sleeman J., Dall P., Ponta H., Hofmann M.;
RT "Characterization of the murine gene encoding the intracellular hyaluronan receptor IHABP";
RL Gene 226:41-50(1999).
RN [4]

RP SEQUENCE OF 164-794 FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=BALB/C; TISSUE=Fibroblast;
RX MEDLINE=96011639; PubMed=7590272;
RA Ewistle J., Zhang S., Yang B., Wong C., Li Q., Hall C.L., A.J.,
RA Movat M., Greenberg A.H., Turley E.A.;
RT "Characterization of the murine gene encoding the hyaluronan receptor RHAMM";
RT Gene 163:233-238(1995).
RN [5]
RP SEQUENCE OF 318-794 FROM N.A., AND CHARACTERIZATION.
RC STRAIN=BALB/C;
RX MEDLINE=92299690; PubMed=1376732;
RA Hardwick C., Hoare K., Owens R., Holm H.P., Hook M., Moore D.,
RA Crilps V., Austen L., Nance D.W., Turley E.A.;
RT "Molecular cloning of a novel hyaluronan receptor that mediates tumor cell motility";
RT J. Cell Biol. 117:1343-1350(1992).
RN [6]
RP CHARACTERIZATION.
RX MEDLINE=94308286; PubMed=7518470;
RA Hall C.L., Wang C., Lange L.A., Turley E.A.;
RT "Hyaluronan and the hyaluronan receptor RHAMM promote focal adhesion turnover and transient tyrosine kinase activity";
RT J. Cell Biol. 126:575-586(1994).
RN [7]
RP ERK REGULATION, AND SUBCELLULAR LOCATION.
RX MEDLINE=98225222; PubMed=9556628;
RA Zhang S., Chang M.C., Zyika D., Turley S., Harrison R., Turley E.A.;
RT "The hyaluronan receptor RHAMM regulates extracellular-regulated kinase";
RT J. Biol. Chem. 273:11342-11348(1998).
RN [8]
RP REVIEW.
RX MEDLINE=99059494; PubMed=9845361;
RA Hofmann M., Assmann V., Fieber C., Sleeman J.P., Moll J., Ponta H., Hart I.R., Herrlich P.;
RT "Problems with RHAMM: a new link between surface adhesion and oncogenesis?";
RT Cell 95:591-592(1998).
RN [9]
RP FUNCTION: INVOLVED IN CELL MOTILITY. WHEN HYALURONAN BINDS TO HMMR, THE PHOSPHORYLATION OF A NUMBER OF PROTEINS, INCLUDING THE FOCAL ADHESION KINASE OCCURS. MAY ALSO BE INVOLVED IN CELLULAR TRANSFORMATION AND METASTASIS FORMATION, AND IN REGULATING EXTRACELLULAR-REGULATED KINASE (ERK) ACTIVITY.
CC -1- SUBUNIT: SUBUNIT OF THE HARC COMPLEX.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE AND CYTOPLASMIC.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: RHAMM1V4 (SHOWN HERE) AND RHAMM1. ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBICITOUSLY EXPRESSED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF031932; AAC12655.1; -
DR EMBL; AF079222; AAD08670.1; -
DR EMBL; AJ005919; CAA06768.1; -
DR EMBL; AJ005920; CAA06768.1; JOINED.
DR EMBL; AJ005921; CAA06768.1; JOINED.
DR EMBL; AJ005922; CAA06768.1; JOINED.
DR EMBL; AJ005923; CAA06768.1; JOINED.
DR EMBL; AJ005924; CAA06768.1; JOINED.
DR EMBL; X64550; CAA45849.1; -
DR EMBL; X64550; CAA45848.1; -
DR MGI; MGI:104667; Hmnr.
KW Hyaluronan acid; Alternative splicing; Repeat; Glycoprotein.
FT DOMAIN 442 546 5 X 21 AA NEAR PERFECT TANDEM REPEATS.
FT REPEAT 442 462 1.
FT REPEAT 463 483 2.

```

FT REPEAT 484 504 3.
FT REPEAT 505 525 4.
FT REPEAT 526 546 5.
FT DOMAIN 719 729
FT CAROXYD 741 750
FT CAROXYD 53
FT CAROXYD 134 134
FT CAROXYD 279 279
FT CAROXYD 446 446
FT CAROXYD 467 467
FT CAROXYD 488 488
FT CAROXYD 509 509
FT CAROXYD 530 530
FT CAROXYD 561 561
FT CAROXYD 601 601
FT VARSPIC 218 242
FT CONFLICT 19
FT CONFLICT 55
FT CONFLICT 71
FT CONFLICT 89
FT CONFLICT 94
FT CONFLICT 540
FT CONFLICT 668
SQ SEQUENCE 794 AA: 91799 MW: 74DB3D36224499C CRC64:

```

Query Match 5.6% Score 223; DB 1; Length 794;

Best Local Similarity 20.1% Pred. No. 0.032; Indels 234; Gaps 40;

Matches 162; Conservative 155; Mismatches 255;

```

OY 45 KEKKNKDNKRKDEE-----RNKTOEHLKMKHYIKVKGGEAYKKE---AAEK 94
DB 76 KSGQNDKRYKLEKRIKRLALOBGRQDKRIQDMSELEKTEKLAANAAREKISLSASNA 135
OY 95 LLEKVPDYLEMYKAIGKIYVDGDTKHSLEALSSEDKKIKIDYGDALLHEHYVA 154
DB 136 SLEKRLTELTRANELLKAK-FSEDC--HOKNMRLALBELMKLRN---KRETKRSMYAK 188
OY 155 KSGYEVLYIOSS-E-DYVENTEKALNVYIEGKIISRLSLKINQ--PQKPLDYLVNTIK 211
DB 189 QSGME--LKLQATQKDLTSEKSKIYQLE--GKLVSE-KEKIDECCEKLELEYIOETS 242
OY 212 NASDSKQDLTLEKHEPTDFSVFE-----LEON---SNEOVFAKAFAYTEPOH 262
DB 243 CASDQVEKKNVIAQLEEDLEKREILSLKQSELEINFFSQIEDLYVKC----- 293
OY 263 RYVOLIYAPAFATNYMDKNEOEINISLE-ELKDQRLMSREKYEKTIKOHYQHSWSLSEE 321
DB 294 ---OLETERDNLVSKDEREAFETLSAEQOILTERLALEROEYEKLOQ----- 337
OY 322 GGLGLKRLPIEPKKDDIHSLSOEKELRLKRIODSSDF---LSTEKEFLKLIQIOI 378
DB 338 ---KELO-----SQSLQOEKELSAKLOOOLCSFOEEMSEKKNVEEELKIAL 382
OY 379 RD-----SLSEEEK---ELNLRIQVSSNPSEKEKEFLKLIQIDOPY 419
DB 383 AELDAVQOKEEESERLVQLEEKSTAEQLTRL---DNLREKEVELEKHAIAAQA 438
OY 420 ---DINRLQDTGLDPSIINDYRKQYKRDIONIDALLHOSIGSTLYNKIYLV 471
DB 439 LIAOEKYNPTAOSLRVYTAQLES-----VOEKYNDPTAOSL---RDVYTAQESQEKY 487
OY 472 EN--NNINNLATGLADLVSDTNKINGINFEKKNKRYSTISNY-----MIVIN-E 523
DB 488 NPTAOSLRVYTAQLESQEKYNDPTAOSLRVYTAQLE---SVQEKYNDPTAOSLRVYSAO 542
OY 524 RPALENERLKRRIQSLPDRAGYLENGKRLIQRNIGLEIKDVOIKOSEKEYIRIDAKVY 583
DB 543 LESYKSTIKLEIEDLK-----LEN--LTLOEKVAMAKSVEDVQO----- 580
OY 584 PYSKIDTKIOEQALNINDEMNKALGLPKYTKLITFVNHNRVYASNIVESAYVLINENKNI 643
DB 581 ---QILTAE-STNOE-----YARWQ-DLQNR-----STL 605

```

```

OY 644 QSDLLKVTNYLVDCNGRVFTDITLPNIAEQYTHODETY-EGVHSGLYVESRSIL-- 700
DB 606 KEKEIKETISSFLE-----KITDLKNOIRQODEDFRQLEKGRRTAEKEVNMKE 655
OY 701 -----LHGPSKGYELRND-----EGFTHERGHAIV-----DDYAGYLLD 734
DB 656 LIMEIKMKRLVYELYEKTRKPPQOQDAFEAEKQALLNENGAIQEOQLNKIRDSYAGLLGH 715
OY 735 KNSQSDLVYNSKRFIDFKREGSNLTS 760
DB 716 QNLKOKI-----KHVYKLKDNSQLKS 737

```

RESULT 7

CENE_HUMAN

ID CENE_HUMAN STANDARD; PRT; 2663 AA.

AC 002224;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE CENTROMERIC PROTEIN E (CENP-E PROTEIN).

GN CENPE.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93024922; PubMed=1406971;

RA Yen T.J., Li G., Schar B.T., Szilak I., Cleveland D.W.;

RT "CENP-E is a putative kinetochore motor that accumulates just before

RT mitosis.";

RL Nature 359:536-539(1992).

RN [2]

RP CHARACTERIZATION.

RX MEDLINE=95196755; PubMed=7889940;

RA Thromer D.A., Jordan M.A., Schar B.T., Yen T.J., Wilson L.;

RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed

RT microtubule motor.";

RL EMBO J. 14:918-926(1995).

RN [3]

RP CHARACTERIZATION.

RX MEDLINE=98437347; PubMed=9763420;

RA Chan G.K.T., Schar B.T., Yen T.J.;

RT "Characterization of the kinetochore binding domain of CENP-E reveals

RT interactions with the kinetochore proteins CENP-F and hBUBR1.";

RL J. Cell Biol. 143:49-63(1998).

CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE

CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE

CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT

CC AND/OR SPINDLE ELONGATION.

CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.

CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING

CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE. AND IS

CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.

CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.

CC

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>

CC or send an email to license@isb-sdb.ch).

CC

DR EMBL: 215005; CAA78727.1; -

DR PIR: S28261; S28261.

DR HSP: P03069; I21T.

DR MIM: 117143; -

DR InterPro: IPR001752; kinesin.

DR Pfam: PF00225; kinesin; 1.

DR PRINTS: PRO0380; KINESINHEAVY.

OY	335	PKKDDIHTSHLSQEEKELKKRQIDSSDF-----LSTEE---KEFLAKQIDIRSLSSE	385
OY	366	EKKELNRIQVDSSHPLESEKER---EFLLKRLKLIDIPYD---INQRKOD-----	427
Db	597	AERAAEVLVMTYHHN---EEKEELTEEMKKLKRRNSNDENEINOITREFFLASNPFKILI	653
OY	428	-----TGLIDSPSIN--LDVRKYQKRDIQNID-----ALHDSIG--STLY	465
Db	654	VTDMLTLGG-FDAPRLKLVMYLDKRPGLYGHRLQAIAARTNRPYPDPKEGFLVDSVGLFVKYL	711
OY	466	NKIYLENNMINNLATLGADLYDSDTNKIKNGICNERK-----KNFKYSISSN	515
Db	712	ETVALYMMLAEEEEREDFEKFNLLSIDSE-----IFGEFKLLEMYKESLSMKLK--INDE	763
OY	516	YMVIDINERPAL-----DNERLKWRIQLSPDRAGUYLNGK-----LILO	555
Db	764	DLSIDVWFTLKTTRKKNKDENNELKEKX---DLIATAYADGGKNAIKLLIDLDAKAVIKLY	819
OY	556	RNIIG-----LEIKDVQIIKQSEKEYIRIDAKVPFSKIDTFKIOEA-----OL	597
Db	820	KALGSYPOKFIFYIEDIELSFI---YAILKKRLPKKSNRRFPWELLISFTINKMLVYDL	876
OY	598	NINQENKALGLPFRYTKLTFFNVNHRASYVESAYLLINEMKNNT--QSLLIKKVTVY	654
Db	877	TVEIEIN-LNPDDLILKENIGCKREIKRAVANFYFIL---KNSIIDQHDPYRK---	927
OY	655	LVDNGRFVFTDITLPRIAE--QVTNODEIYEGVHSKGLYVESRSILLHGSKGVETLRN	712
Db	928	-----ELIERLERLRDMWTMKNRIDK-----TYNAIKNLMELN	962
OY	713	D-----SEGTHIEFGHAVDVYAGYLLDKNQSDL--VTNSSKFIDFEKESGNLTSYGR	763
Db	963	NYDKIKIGKSIERIKSIISTYIGENILKQD-DIKNLNLENTELK----TKQONNLKLSK	1019
OY	764	TNEAEFFAE 772	
Db	1018	LQRRKFFKE 1026	
RESULT	9		
ID	MSJ_DICDI	STANDARD;	PRT; 2245 AA.
AC	P54697;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	MYOSIN IJ HEAVY CHAIN.		
GN	MYOJ.		
OC	Dictpostelium discoidem (Silme mold).		
OX	Eukaryota; Mycetozoa; Dictpostellida; Dictpostelium.		
NCBI_TaxID	44689;		
ZN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-AX3;		
RX	MEDLINE=96215148; PubMed=8636147;		
RA	Hammer J.A., III, Jung G.;		
RT	"The sequence of the dictpostelium myo J heavy chain gene predicts a		
RT	novel, dimeric, unconventional myosin with a heavy chain molecular		
RT	mass of 258 kDa."		
RL	J. Biol. Chem. 271:7120-7127(1996).		
RN	[2]		
RP	SEQUENCE OF 1-1021 FROM N.A.		
RX	MEDLINE=97039016; PubMed=8864597;		
RA	Peterson M.D., Uristote A.S., Titus M.A.;		
RT	"Dictpostelium discoidem myoJ: a member of a broadly defined myosin		
RT	v class or a class XI unconventional myosin";		
RL	J. Muscle Res. Cell Motil. 17:411-424(1996).		
RN	[3]		
RP	SEQUENCE OF 182-298 FROM N.A.		
RX	MEDLINE=95023928; PubMed=7937787;		
RA	Titus M.A., Kuspa A., Loomis W.F.;		

```

RT "Discovery of myosin genes by physical mapping in Dictyostelium.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9446-9450(1994).
CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN AND HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL: U42409; AAA85186.1; -.
DR EMBL: L35322; AAA79858.1; -.
DR HSP: P08799; 1MND.
DR DictyDB: DD01095; myoJ.
DR InterPro: IP002710; myoJ.
DR InterPro: IPR000048; IQ.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF01843; DIL; 1.
DR Pfam: PF00612; IQ; 3.
DR Pfam: PF00063; myosin_head; 2.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRODOM: PD000355; myosin_head; 1.
DR PRODOM: PD003376; DIL; 1.
DR SMART: SM00015; IQ; 3.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 3.
KW Myosin; ATP-binding; Actin-binding; Calmodulin-binding; Repeat;
KW Coiled coil.
FT DOMAIN 1 809 MYOSIN HEAD-LIKE.
FT DOMAIN 824 851 IO 1.
FT DOMAIN 872 901 IO 2.
FT DOMAIN 943 972 IO 3.
FT DOMAIN 973 1812 COILED COIL (POTENTIAL).
FT DOMAIN 1813 2245 TAIL.
FT NP_BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 669 749 ACTIN-BINDING.
FT CONFLICT 191 191 L -> F (IN REF. 2).
FT CONFLICT 284 284 A -> T (IN REF. 2).
FT CONFLICT 291 291 G -> R (IN REF. 2).
FT CONFLICT 332 347 NKGGEFEEGVSDDEH -> IEMFELKVRNKS (IN
FT REF. 2).
FT CONFLICT 550 550 N -> K (IN REF. 2).
FT CONFLICT 865 866 HH -> QQ (IN REF. 2).
FT SEQUENCE 2245 AA; 258478 MW; 615E5EFLIDAB4SBE CRC64;

Query Match 5.4%; Score 215; DB 1; Length 2245;
Best Local Similarity 19.4%; Pred. No. 0.22;
Matches 169; Conservative 147; Mismatches 274; Indels 280; Gaps 40;

QY 50 NKDEKKRKDEER-----NKQEE---HLKELMKHYV--IEVGGE-- 85
DB 1123 NKOEDFRLSQEDRDTNTNNQLLEIOLKRNASTYLEEDFSLGIRNLEKROVELLELDENOL 1182
QY 86 -----AVKKEAEKLE-----KVPSDVLEMYKAIGKITYVG 119
DB 1183 IKERLDSLGQSSSQFQSGAALFKQOLEVLVQDSQEDLIKLSSEKLGSEEAQKQINOLEL 1242
QY 120 DITKH-----ISLEALSDEKKRKIKDKYKDALLHEHYVYAKGEYEPVLVIQSSDEVYNT 174
DB 1243 ELTDHKSRLQIQLOLQTEGSENEKIKTKLKG--LEFYQDEKQKQOLERIRKQSGSVDE 1299
QY 175 EKALNVEYEIGKILSRDLISKINQYQKRLVDLNTITKASDSDG----- 218
DB 1300 KNSLTTLQTVVFEETQVSTNVSHQEKITTKTKSTLEELNKSIGKIQLOAEQNKKDEIRKI 1359

```

```
QY 219 -----ODLFTNOLKEH-----PIDESVEFLQNSNEVOEFAKAFAYIEP 260
      |||||
Db 1360 QFELNDKQOFTKQTEFSLQSQSDIDRPKSETTHSHLERTNETLKSDEERY-QQSLKQ 1418
      |||||
QY 261 QHRDVLQLYAPEAFNYWYDKFN--EOEINLSLEELKD-----294
      |||||
Db 1419 QHRDCCQ-----YKDTJNRLNEVK-QLTQKKEFNEFFAKQNSNQOESYTK 1469
      |||||
QY 295 -----ORMLSRREK-WEKIKOHQWSDLSSEGRGLKLQIPIPKKDIITHLSOE 347
      |||||
Db 1470 EYTTQMOQNSRIERELEERKOHITRIDE-RDELKQLTOLQOQHOSSTOLL--LAON 1526
      |||||
QY 348 EKELKRIOT--DSSDFLSTEEKEF--LKKIOLDIRDSLS-----EEKELNRIQVD 396
      |||||
Db 1527 ELERLKKELKYERGHETSKQOQFMETIOSLKITNNDLQSLQDYOEKKKL-KDKLS 1585
      |||||
QY 397 SSNPSEKEFEFLKLKLQIOPY-----DINORLQDTGGLIDPSINLADVRKQ 444
      |||||
Db 1586 SSKQEQOQRESIHKMAELSAIKQHSQWVENSFTDMKKQNOE--LIESSAL--YKQ 1639
      |||||
QY 445 YKRDIONDAL-----LHQ-----SIGSTLYNK--I 468
      |||||
Db 1640 LLOOTSTIDSTIREKENESIKLOOLETSNQOLHOLKEELNSMKOSQOLESTEQSKOLNQ 1699
      |||||
QY 469 YLYENNNINLTATIGADLVDSF--DNKIKNGIFNEKKKPKYSISSNY-----516
      |||||
Db 1700 LIGENQOLKSVTEISKQOLDVAFENOKINNTKEOEIKRMSVVELOQHIDEGKQOEIQ 1759
      |||||
QY 517 -----MYDIVNERPALDNERLKWRIQ-----LSPDTRAGY--LENGKLLIQRNIG 559
      |||||
Db 1760 QLOSTIAQLQOQOSESFDRLKEKEIQMKRERETQMKLVESTKLNLYHLED-RMELYKNV- 1817
      |||||
QY 560 LEKDVQITQSEKEYI-----RIDAKVPPKSIDTKIQEAOININQEW-----603
      |||||
Db 1818 MEIID--YKETEMEKARLAGCKELDTKLISDFLSCKEHTSLG--SQWMFHQIDWCP 1873
      |||||
QY 604 -----NKALGLPYTKLTITFNVHN--RYASNIVESAVYLINMKKNISDILKKATNY 654
      |||||
Db 1874 YERDSSKGIYGIIRSIYDVTIKNFEDVDLSTYLACCSTLFLYKKNLYKHL-----1926
      |||||
QY 655 LVDNGRFEFTDI-----TLPNIAEOYTHQ 679
      |||||
Db 1927 -----NGANSIMPIIPTGLDELEINERLSHQ 1952
      |||||

RESULT 10
YD86_SCHPO
ID YD86_SCHPO STANDARD; PRT; 1957 AA.
AC 010411;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOHETICAL 222.8 KDA PROTEIN C1F3.06C IN CHROMOSOME I.
GN SPAC1F3.06C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
OX (1)
RN SEQUENCE FROM N.A.
RC STRAIN=972;
RC Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
RL
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL_270690; CA94624.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

Query Match 5 3%; Score 209.5; DB 1; Length 1957;
Best Local Similarity 18.9%; Pred. No.0.31;
Matches 186; Conservative 131; Mismatches 330; Indels 317; Gaps 44;

QY 59 EERNKQOEELHKEIMKHIVKVEKGEEAVRK-----EAAEKLEKVPDSEVLEMYKAIG-GR 113
      |||||
Db 650 EOKESTLSDENNDLRTFLKLEESNKSILKQEDVDSLEKNIDQTLKEDDKLSEALRFK 709
      |||||
QY 114 I-----YIYDGDITKHISLEALSED-KKKIKDLYGDALLHGHYAAKGYEPV-----161
      |||||
Db 710 LEAKNLRREVINDLKGKHELEAQRNDLHSSLSPAKNTNALISELTKSSSDVRLTANVE 769
      |||||
QY 162 -----LVYQSDSEYVENKEKALNYYE-----183
      |||||
Db 770 TLQODSKAMKQOSTSLVNSYQISINLYHELRDHVNQNSQNTLLESESKLKTDCENLQ 829
      |||||
QY 184 -----IGKILSRDI-----LSKINQPYQKFLDYVNTIKNASDSDGODLFTNOL 227
      |||||
Db 830 QNMTLIDNVQKLMKHKNHNOESKVSSELKEVNGKLSLDLKNLRSSLNVAISDNDQL--TQL 887
      |||||
QY 228 KEHPDSEVEFLQNSNEVOEFAKAFAYIIEPOHRDVLQLYAPEAFNYWYDKF-----280
      |||||
Db 888 AELSKNY--DSLQESQOLNSGLKSLEA-----EKOLLTFENEELHIRLDKLTGKLTIE 939
      |||||
QY 281 -----NEORI--NLSLSELKDOMLSRYE-----KMEKIKOHQWMSDSL 318
      |||||
Db 940 ESKSSDLGKKLTARQEBISLKEENMSQOAITSVKSKLDELTSKSKSLADIEHLKNKY 999
      |||||
QY 319 SE--EGRGLL-----KKLQIPIEPK--DDIHSL-----344
      |||||
Db 1000 SEVEVERNALLASNERLMDLKNNGENIASLQTEIERKRAENDLQSKLSVSEYENLL 1059
      |||||
QY 345 ---SQEKEEL--LKRQ-----IDSSDFLSTEEKEFLKL-----374
      |||||
Db 1060 LISSQTKSLSDKTNOIKYTEKNVQKLLDEKQDQANVELLELTJRYKGLGEMNOIDELL 1119
      |||||
QY 375 -----OIDIRDSLEEEKELNRIQVDSN--PLSEKEFEFLK 410
      |||||
Db 1120 ALRKKSKKQHDLCANFPVDLKEKSDALEFQLTNKEKNELIYLSQNSNENALYVERSDLAN 1179
      |||||
QY 411 KKLQIDQYDINQRLQDTGGLI-----DSPSIN--LDVRK-----QYKRDIONDAL 455
      |||||
Db 1180 RLS-----DMKKSLSDDSNVIVISYRSDLYRVNDELDTLTKDKDSLSTQYSEVQQRDLD 1233
      |||||
QY 456 LHSIG-STLYNK--IYLYENMMINLTATIGADLVDSVTNTKINGIFNEPK-----506
      |||||
Db 1234 LDSLKGCESEFNKYAVSLRELCTKSELIDPV-SEIID--DNFVFNNGNFSELSRLTVLSL 1290
      |||||
QY 507 -----NFKYSISSNNY-----IYDINERPALDNERLKWRIQSPDTRAGY 546
      |||||
Db 1291 ENYLDAFNQVNFKMKELMDNLTTTDAFTKVVAADLER--LQHEHDMILQ-----RGD 1341
      |||||
QY 547 LENGKLLQRNIGLEIKDVQITQSEKEYITRIDAKV-----VPKSIDTKIQEAOININ 600
      |||||
Db 1342 LE-----KALKDEKKNFLRKAEMTENTHISLEEGKEETKEIAELSSR 1384
      |||||
QY 601 QENNKALGLPKYTKLTITFNVHNRYASNIY--ESAVYLINEMKNN--IOSDLI--KKVT 652
      |||||
Db 1385 LEONQALTNKLNQDLHNLQRIKEDVLKEKESLIITISLESLSNOROKESSLLDAKNEL 1444
      |||||
QY 653 NYLVNGRFEFTDITLPNIAEQYTH--ODEYE-----OVHSGKGYEPESRS 698
      |||||
Db 1445 EHMIDTDR---KNSLMEKIESINSSLDKSEFLASAVKALQKLHSESLMENIK 1501
      |||||
QY 699 ILLHGPSGYELRNDSEGFITHEFGHAV---DDYAGYLLDKNO--SDLYTNSKRFIDIFK 752
      |||||
Db 1502 SQLQEAKEKIQV---DESTIQELDHEITASKNNYVEGKLNKDSIIRDSINIQNLNLLA 1558
      |||||
```

QY 753 EEGSNLTSYGRTEAEFAEFAEFL 776
DB 1559 EKSNAVKRLSTESKEIILPNSRL 1582

RESULT 11
Y109_YEAST STANDARD: PRT: 1679 AA.

AC 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 01-FEB-1995 (Rel. 31, Last annotation update)
GN HYPOTHETICAL 195.1 KDA PROTEIN IN DNA43-UB11 INTERGENIC REGION.
Y11449C.

OS Saccharomyces cerevisiae (Baker's yeast)
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.

OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Barrill B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Hornell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajadream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: 238059; CAA6129.1; -
DR PIR: S48385; S48385.
DR SGD: S0001411; MLP2.
KW Hypothetical protein.
SQ SEQUENCE 1679 AA; 195141 MW; 298950CC52202D8F CRC64;

Query Match 5.2%; Score 209; DB 1; Length 1679;
Best Local Similarity 19.8%; Pred. No. 0.27;
Matches 171; Conservative 155; Mismatches 272; Indels 264; Gaps 42;

QY 44 VKEEKNKDENRKDEER---NKTQEHLEIKMHIKIEVGEBA---VKKEA 92
DB 644 IYKSGCKKKTLEDPENKGLAKEKRMLEALDHL-KAELEKQSWPSYTHVEKRA 702
QY 93 EKLEKVPDVLWKYKAIGKTIYVDGDTKHISLEALSDKKKIND---LYGDA 147
DB 703 STELSQ-----SRIKISLEYEISLEKKTETASPIPTKESLT 738
QY 148 HEHYVAKGEYPLVYIOSSSEYVENTEKALNYYEIGILSR-----DILSINPYQK 202
DB 739 RPEQCCCKKKELQMRLEKE---ISHNENKMPFSSKEGGYKAKIKELNLELRSLDLOS 796
QY 203 PLVDLNTINASDS-----DGODLLFTNOLKE-----HP 211
DB 797 KIOEIESISCKDSOLKAMQNTIDTEMKMSILTELKSKETTIEKLSIEINLDELRK 856
QY 232 TDFSEVELEONSNEVOEYAKAFAYIEPQHR---DVLOLYAPEANMYDKFNE--QETIN 286
DB 857 TFEQKFLQNSD-----ASTLEPTLRKELEQIOVOLKANDASQIOAVEIISNNE 906
QY 287 LLELEKQDM-----LSREKWEKIKQHYQHSNDSISEGRLGLKLQIPIEP 335
DB 907 NLILEKELATKKNYDAKIELEKKEKWAR-----EEDLSRL-RGELGIR-ALQP 956
QY 336 K-KDIIHLSLOEEK--ELLKRIQ--IDSSDFLSTEEKFLKLQIDIRDSISEEK--- 387

DB 957 KKEGALHFVQOSEKRLNERIQRKMIETKEKST--IVQCKKKEKSOYSTKKEKDDLS 1015
QY 388 ELLNRIQVDSNPLSEKEFEFLKLKDIOPYDINORDPTGLDPSFIN-LDVKKQYK 446
DB 1016 ELVIRLEKDAADQAE-----LTKTK-----SLSVSAQDLIDKHEKWEKAE 1061
QY 447 RD-IQNDALHOSIGSTLYENKINNNLTITGLADVLDSIDNKRINKINGINERK 505
DB 1062 RELISNIEQT-----ESLRVEN-----SVLLEKVDTPAANNDDCHLK 1099
QY 506 KNEFYSSNMYMIDINERPALDNERLKRIOQSPPTFRAGYLENGKLIILORNIGLE--TK 563
DB 1100 ---LVSLFNLR-----HERSLETKLTCKRELA-----FVQOKNSLEKTIN 1140
QY 564 DVOIIO-SEKEYRIDAKVPSKIDTYIQEAQLINQINQEWKALPLPYTKLITFNVN 622
DB 1141 DLORTQTLSEKEY-QCSAVIIDEFKDITK-EYQVNIILKE-NNAI-LQSKLVTEKNE 1196
QY 623 RVA-----SNIV-----ESAYVLINENKKNIOGD 646
DB 1197 IYKQLNDROEISRLQDRLIQTKQVYSINSKILVYSEMEQCKORYODLSQOKDAOKK 1256
QY 647 LIKKVTNVLVDGNGRFRFTDITLPNIAEQYTHODEIYGVHVK-----689
DB 1257 DIEKLTNEISDLKGKLSAENANADLENK---NRLKQAKHEKLDASKKQOALTNELNE 1313
QY 690 -----GLVPEPSRSTLHGPPSKVGLRNDSEGFTHFEGHAVDVAAGYLLDRNQSD 739
DB 1314 LKAIKKLEQDLHFENAKVIDLDTKLKAHLQSEDEVSRDHE---KDTY-----RTLME 1363
QY 740 LVYNSKKFLDIFKEEGSNLTSY 761
DB 1364 EIESLKEQIIFKTANSSDAF 1385

RESULT 12
Y104_YEAST STANDARD: PRT: 691 AA.

AC 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
GN HYPOTHETICAL 80.5 KDA PROTEIN IN SLN1-RA25 INTERGENIC REGION.
Y11444W.

OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.

OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Barrill B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Hornell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajadream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: 238059; CAA6134.1; -
DR PIR: S48390; S48390.
DR SGD: S0001406; TID3.
KW Hypothetical protein.

SO SEQUENCE 691 AA; 80487 MM; 7DBC492227A80093 CRC64;

Query Match 5.2%; Score 208.5; DB 1; Length 691;
Best Local Similarity 20.5%; Pred. No. 0.1;
Matches 132; Conservative 117; Mismatches 195; Indels 201; Gaps 32;

```

OY 119 GDTRKHSLSLSEDKKKIKIYKGDALLHEHYAK-----EGEYV-----161
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 105 GSLQHLST-----NRDPRLRDKNFQSAIQEETIYDLKKNKFDIEIHHISIKFLKOPTOK 160
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 162 -----LVIOSSDYENTKALNVYEIGKILSRILEKINP-----YQKFL 204
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 161 GFTITFKWLRLDPGY-GFTKSIENEIYQLKRLRYPELISINSQISAVGSGNMHFL 219
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 205 DVL-----NT-----IKNASDSDQ-----DLFTNQL 227
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 220 GMLHMVATNIKLMCLNKVDRLINQTOETILSLOPLKLDLDDQDQREHYELMVERLL 279
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 228 KEHPTDSVER--LEQNSNEVQEVFAKFAFYIEPOHRDVLQLYAPAFNYMDKFNBOEI 285
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 280 IDYTESYKSKFLLEDNEPSEMELKLEFEKFEVHIINTDINLOT-QDNLYERY--QEV 336
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 286 NLSEELKDQMLRYEKWEKIK--OHYQMSDSL--SEEGRLKLLQIDIEPKKD 339
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 337 MKISOKITTR-----EKMKALKSDSKNYENYVAMAKOSQEMFCKLEKMSCELEKEE 391
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 340 TIHSLQSEKELKRIQIDSDPLSTEEKEFLKLLQIDIRSLSEEEKELNRIQVDSN 399
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 392 -IKALQSNISELHKILR---KKGISTEOLF-----LQNGEREKLTR-ELDKIN 434
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 400 PLSEKEKEFLKLLDQPYDINQRLDTGGLIDSPSINLQVROYKRDIONI-----452
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 435 IQSKLTSSISKRLKEV-----GIEFS--LDLTQLQYSSISQNLRSRSQL 479
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 453 -----DALHOSIGSTLYNKIYLENNINMLTATGADLVSDTDMT--KINRGIFNEFK 505
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 480 GHNVNDSLKINISENLDRDF-HGISYBQLPFR-GSGINSEIKSLIKLDELQOEIK 537
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 506 -----KNFYSSISNMYIVDINERPALDNERLKWRIQLSPDRAGYLENGKL 552
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 538 TIERKDNITLEKIDINKLKHDINEK--TQINKELELESEANSKPELSKO-----ENRLL 588
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 553 ILQNGIIEIDVOIIOSEKEIYRIDAKVYPRSKI-----DKIOEADQNTINQOE 602
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 589 LVAQRIELEKEKKI-----NDSNLLMKTRISDAEELVSTELKLELVVDLNRK 638
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 603 WNKALGLPKYTKLITFVNHRYASNIYESAVLLINEMKNIIQSDI 647
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 639 -----RYKLHQQ-VIHHVID-----ITSKFRINIQSSSL 664
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 13
SMC3_YEAST STANDARD; PRT; 1230 AA.

```

AC 01-FEB-1996 (Rel. 33, Created)
AC 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CHROMOSOME SEGREGATION PROTEIN SMC3 (DA-BOX PROTEIN SMC3).
GN SMC3 OR YJL074C OR J1049.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OC NCBI_TaxID=4932;
OY NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M303;
RX MEDLINE=97474309; Pubmed=9335333;
RA Michaelis C., Cloak R., Nasmyth K.;
RT "Chesins: chromosomal proteins that prevent premature separation of
  sister chromatids";
RL Cell 91:35-45(1997).

```

RN [2]
RA SEQUENCE FROM N.A.
RA Rose M., Koeltter P., Entian K.D.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA Sor F.J.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC PART OF A CHROMOSOME CONDENSATION MOTOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL). COILED-COIL REGIONS
CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Y14278; CA474655.1; -;
DR EMBL: 249349; CA89366.1; -;
DR EMBL: X88851; CA61313.1; -;
DR SGD: S0003610; SMC3.
DR InterPro: IPR003439; ABC_Transport.
DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02483; SMC_C; 1.
DR Pfam: PF02463; SMC_N; 1.
KW Mitosis; ATP-binding; Coiled coil; Nuclear protein.
FT NP_BIND 32 39 ATP (POTENTIAL).
FT DOMAIN 172 482 COILED COIL (POTENTIAL).
FT DOMAIN 685 1041 COILED COIL (POTENTIAL).
SQ SEQUENCE 1230 AA; 141336 MM; B152D86F7780341F CRC64;

Query Match 5.2%; Score 208; DB 1; Length 1230;
Best Local Similarity 20.1%; Pred. No. 0.21;
Matches 162; Conservative 146; Mismatches 256; Indels 242; Gaps 39;

```

OY 57 KDEERNKTOEELKELKMKHYIEVKEEA--VKKEAEKLEKVPDVLMEYKAGKI 114
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 158 KDKERLQLE-----DVVGAKSFEVYKLKASLKMEETQKKIQINKEMG---201
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 115 YIVGDITKHISLSELDKDKKIK--DIYKDALLHEHYVAKGEYV-----161
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 202 -----ELNSKLSLSEMDQERKELKYNELERNKRYOFTLYDRELNEVINOMERLDGYN 254
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 162 LVIOSSDYENTKALNVYEIGKILSRDILSKINOPYKFLDVLNTIKNASDSDQDL 221
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 255 NTVYSSEGIQELDKREDMDQVSKLSS-----IEASLKINKATDLO-QAK 300
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 222 LFTNQLKEHPTDSVEYLEQNSNEVQEVFAKFAFYIEPOHRDVLQLYAPAFNYMDKFN 281
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 301 LRESEISQKLTVNWKI-----KDVQO-----QIESNEEQRLDSAT 337
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 282 EOEINLSLEELKDQ--NMLSYREKWEKIKOHYQMSDSLSEEGRLKLL--LQIDIEP 335
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 338 LKEIKSITTEOKKOKSLKRLPYQELTYEEAMYKQLASLQOKDILLIKGEYARFKSKD 397
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 336 KKKDIISLQSEK-----ELKRIQIDSDPLSTEEKEFLKLLQIDIRDSLSEEEK 388
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 398 EROTWIHSEIEELKSSIQNLNLESQLOMDRTSL--RKQY-----SAIDEEIE 444
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 389 LNLRIQ--VDSSNPLSEKEKEFLKLLQIDQPYDINQRLDTGGLIDSPSINLQVROYKR 447
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 445 LIDSINSPDTGQLDEPDSLDI-----HLKQKLSE-----SLDTRKELMR 484
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 448 DIQNDALHOSIGSTLYNKIYLENNINMLTATGADLVDS-----TDNFKINR--498
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```



```
Db 485 KEOKL-----QTVLETLTLDV-----NQNRNVNFTMSRSLANGIINVKEITEKLKISPS 535
Qy 499 --GTFNE-FKKNFKS-----ISSN---YMYDINERPAL-DNE-RLK-WAI-----Q 537
Db 536 VFTGLGELIKYNDKRYKCAEYIGNSLFIYVDTEATLLINNELYRKGKGVTEPIPLNR 595
Qy 538 LSPDTRAGYLEN-----GKLLIQRNIGLEIKVQIIR 569
Db 596 LSLSDVAFPSNTTQIOFTPLIKIKYEPREKAVKHFGTTIYVKDIGGLK---LAK 652
Qy 570 QSEKEYIRIDAKVVPKSIDTKIOEAOLINQENKALGLPYTKLITFNHNRYS--- 626
Db 653 KHKLNATILDDGRADKRGVLV-----GGYLDQHKRTLESJLKNLNESSRQHK 699
Qy 627 NIVESAVIILEMKNKNIOSDLIKKYTNLVDSNGRFVFTD--ITLPNTAEQYTHODEIYE 684
Db 700 KILEELDFVRND-----LNDIDTKIDQ--VNGNIRKVSNDRESVLTNI----- 740
Qy 685 OVHSKGLVVPERSRSLILHGPSKGYELR-----NDSEGFIEREHAVDVAGYLLDKNQSDL 740
Db 741 EYRTSLTKTKNEKILTEESLNATILKLEKLTNTNTFAOE--KLNTENDLLOEFDSEL 797
Qy 741 VTNSKRFIDIFKEEGS-----NLTS 760
Db 798 SKEKERLESJLTKELSAHNKLNITS 823

RESULT 14
Y328_MYCGE STANDARD: PRT: 756 AA.
AC 049419: 049308: 049320:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MG328.
GN MG328.
OS Mycoplasma genitalium.
OC Bacteria: Firmicutes: Bacillus/Clostridium group; Mollicutes.
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=756993;
RA Frieser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier J.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 19-113 AND 155-272 FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U39114; AAC71552.1;
CC EMBL; U02203; AAD12492.1;
CC EMBL; U02188; AAD12472.1;
CC DR
```

```
DR TIGR; MG328;
KM Hypothetical protein: Complete proteome.
FT CONFLICT 52 52 T -> R (IN REF. 2).
FT CONFLICT 65 65 S -> W (IN REF. 2).
SQ SEQUENCE 756 AA; 88407 MW; 1D3AD4EFOAB5C068 CRC64;

Query Match 5.2%; Score 205.5; DB 1; Length 756;
Best Local Similarity 22.2%; Pred. No. 0.15; Indels 219; Gaps 42;
Matches 158; Conservative 120; Mismatches 215;

Qy 46 EKEKN-----KDEKRRDE-----ERNTOEHLKEIKHIVKIEVKEEAV 87
Db 61 EKEKIVSEFNMMDLKYDKKDKKLVNHNANOLKTRREBOGYVQGTAFQEVDOQSLV 120
Qy 88 KKEAAEKLEKVPY-----DYLEMYKAIGKIYIVDGTITKHSLEAL-----SED 133
Db 121 TKEIKTGFQELKPSVYTAEDDKDE-----IKPATKQVSPFEELNQPSSEINET 169
Qy 134 KKKIDYKQKALLHEHYAAKEGYEPVVIQSSDEY--VENTERKALNYVEIKGLSRD 191
Db 170 KRPEVQIFSTDRV-----REP--EQFDFYSIENLTKAIN--PVHKTIOYD 211
Qy 192 ILTKINQPY--OKFLDVLNTIKNASPD--GODLLFTN-QLKEHPTDPSVEFLQNSME 245
Db 212 --QNDQPFVVKRILKEQHPTKKVDELDYNNKELLEBNADLK-----QIDDKKNND 264
Qy 246 VOEVFAKAFAYYIEPQHRVLIQYAEAFAYMDKFNQOEINISLEELKDQRMLSRYEKWE 305
Db 265 --QIF-----DL-----EQEI--DDLK--RRLSE---E 283
Qy 306 KIKQ-HYQWSDSLSEEGGLAKIO--IPLEPKKDDIITHLS--QEEKELKRQIDS 359
Db 284 KKHHLTKRLQDLDLQENDLYEQLNKPVALNPISDEVNEELNKKQKALLSD-QLDA 342
Qy 360 SDFLSTEEKEFLKLIQIDIRDSISEEKEELNRIQYDSSNPISSEKEEFLK--KIKLDI 416
Db 343 LNKSSNVQOQLALPVR-LNNOINELQNLITAREANQRLDLYEQNDPLKMKELKHLN 401
Qy 417 QPYDINQRLQDTGG-----LIDSPINLDVRKQYKRDIONIDALHOSISTLYNKIYLYE 472
Db 402 TSNDENEKYYDDLNOYELLFEDENETKFD-KIQVQOQALMD--YOKTISALKHE----- 452
Qy 473 NNNINNLTTLGADIVDSDNTKINGIFNEFKKNKYSISSNMYLVDLNERPALD----- 528
Db 453 ---NDV-----LDELIEWTRSKDNDFNNTKNSPE-----EOKKALDEKLN 489
Qy 529 -----NERLKWRI-QLSP-DTRAGYLENGKLLIQRNIGLEIKDVQIKQSEKEYIRIDAK 581
Db 490 GLTIQNOQLQDRIAELEBENNEKSNLNTQIV--NLQOQLKDSQMLFN----- 535
Qy 582 VVPKSIDTKIOEAOLINQENM--KALGLPKYTKLITFNHNRYSNIVESAVLIIN 637
Db 536 -VAQDKLAT-LREVNLAINEKINDLEDELGS-----SENSNNLLAK--LQADHEILQ 583
Qy 638 EKKNNIQSLIKKYTNVLYVDGNGRFVFTITLPNTAEQYTHODEIYEOVHSK 689
Db 584 ESYGKLKTDPERLKKKKLNDAN-----EYQVODLLSAFEETNSE 621

RESULT 15
MLP1_YEAST STANDARD: PRT: 1875 AA.
AC 002455;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN-LIKE PROTEIN MLP1.
GN MLP1 OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RA MEDLINE=93247549; PubMed=8483450;
RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RL Mol. Gen. Genet. 237:359-369(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MPl1 genes and three
RT new open reading frames.";
RL Yeast 9:1349-1354(1993).
CC -1- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC -1- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -1- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPl1".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L01992; AAA34783.1; -
DR EMBL: X73541; CA51948.1; -
DR EMBL: Z28320; CA82174.1; -
DR PIR: S38173; S38173.
DR SGD: S0001803; MPl1.
KM Coiled coil; DNA repair.
FT DOMAIN 69 487 COILED COIL (POTENTIAL).
FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
FT CONFLICT 301 301 R -> A (IN REF. 1).
SQ SEQUENCE 1875 AA; 218455 MW; 683AD034C9066867 CRC64;

```

```

Query Match 5.28; Score 205.5; DB 1; Length 1875;
Best Local Similarity 20.44; Pred. No. 0.42; 291; Indels 269; Gaps 51;
Matches 186; Conservative 164; Mismatches 291; Indels 269; Gaps 51;
QY 2 NIKREFIKVI-----SMSCLVTAI-TLSGPVFIPLVQAGHGVDGMHYKEKKNKKDKR 56
DB 765 NLKQELNKLSPKESDIRMTVQLOTLQ-----KEREDLLEETRK 803
QY 57 KDEERNKTOEHLKEIMK-----HIYKIEVGE-----EAVKKEAEKLEKLV 99
DB 804 SCQKKIDELDELSELKETSQKDHNIKQLEEDNNSNIEMWQNKIEALKKD-YESVITSV 862
QY 100 PSDYLEWYKAGGKIYVGDY-----TKHISLEALSDDKKIKDIYKDALLHEHY 151
DB 863 DSKQTDIEK-LQYVVKSLKEIEEDKIRLHTYVWDEFINDSLR-KELEKSKINTLTDAY 920
QY 152 VYAKEGEYPLVIOSEDEYENTFEKALNVYIEIKIISRLISKINQPYQKFLDLVLTIK 211
DB 921 SQIKE-----YKDLVFTTSQSL-----QQTNSKLD---ESFKDFTNQIK 956
QY 212 NASD-----SDGQDL-----FTNQL-----KEHPTFS--VEPLEONSNEVEVFA 251
DB 957 NLTEKTSLEDKISLKEQMFNLNNEEDLQKKGEKADPKKRISILQNNKKEVAVKS 1016
QY 252 KAFAYYTEPQH-RVQLQYAPEA-FNYWDRKNEQ-EINLSLEELKDQ-----RML 298
DB 1017 EYESKLSKIQNDLQOQTIYANTQANNYEQELQKHADVSKTISELREQLHTYKGOVKTLNL 1076
QY 299 SRYEKMEKIKOHYQHS-----DSLSEGRGLKQLQPIEPKRD--- 338
DB 1077 SRDQLENAKLKENEKSWSSQKESLLEQDLDSNRSRIEDLSSQNKLLYDQIQIYTTAADKEVNN 1136

```

```

QY 339 -----DIHHSLSQEKELKRIQIDSSDFLSTEKEEFLKLOI-DIRDSL--SEEK 387
DB 1137 STNGPGLNLTILTLRRRDLIDTQVTAERDAAKRLRQKISLMDVELQDARTKLDNSRYEK 1196
QY 388 E-----LNKRIQVDSNPL--SEEKEF--LKLKLDIOP--- 418
DB 1197 ENHSSIIQOHDDIMEKLNQNLRLRESNITLRLNELNNNNKKKELQSELDRKQVAPIES 1256
QY 419 -----YDINORLODTGGLIDPSINLDVVRKQ---YKRDIONDALHQSIGSTLYNKI 468
DB 1257 ELTLAKYSMOEKEE-----LKLAKYHRRKKKSQDI-LEKHQQLSSDYEKL 1304
QY 469 YLYENNMNNLTATL-----GADLVSDTNTKINRGIFFNEFKNFYYS-ISSNYMIVDI 521
DB 1305 -----ESEIENLKEELEKKEKRGAEAE-----KFNRLRRQAQERLTKSKLSQDSLTEQV 1354
QY 522 NE-----RPALDNERLAKRIQUS-PDTRAGYLENGKLLQNRNIGLE-IKDVQIIKOSKEK 575
DB 1355 NSLRDAKNVLEN-----SLSEANARIEELQNAK-VAQGNNDLEALRKLO--EDAEKAS 1404
QY 576 IRIDAKVPRKSIDTKIQEAOLN-----INQENKALGLPRYTKLITFENVHNRVASN 627
DB 1405 RELOAKL-----EESTYESTINGLNEEITTLKEIEIKQRIQOQLQATISANEQD-LSN 1459
QY 628 IVESAYVILNEMKNNIQSDILIKKYVNYLVONGRFEV-----TDITLPNTAEQ--T 677
DB 1460 IVES-----MKSPFEDKIKFIKEKTQEVNKEKILBAQERLQPSNINNEEIKKWESE 1512
QY 678 HQDEIYEQVH-----SKGLYVPRSRISILHGSPKGYELNDSGFIHEGHVADVAGY 731
DB 1513 HEDVSVQKIRPAEALAKRIRLPTREKI-----NKITIERK--KEELEKEFEKVEERIKS 1565
QY 732 LLDKNSQDLV 741
DB 1566 MEQSGEIDV 1575

```

Search completed: December 2, 2001, 13:51:46
Job time: 341 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:46:00 : Search time 55.64 Seconds
(without alignments)
327.196 Million cell updates/sec

Title: US-09-747-521-2
Perfect score: 4145
Sequence: 1 MNKKKFIKIVSMSCIVTAI.....KNAPKTFQINDQIKFIINS 809

Scoring table:
BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents_AA:*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/6CTUS.COMB.pep:*
7: /cgn2_6/ptodata/2/1aa/6ackfilesl.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3983	96.1	776	1	US-08-021-601-2 Sequence 2, Appl1
2	3983	96.1	776	1	US-08-082-849B-2 Sequence 2, Appl1
3	3983	96.1	776	5	PCT-US94-01624-2 Sequence 2, Appl1
4	1309	31.6	472	1	US-08-021-601-8 Sequence 8, Appl1
5	1309	31.6	472	1	US-08-082-849B-8 Sequence 8, Appl1
6	1309	31.6	472	5	PCT-US94-01624-8 Sequence 8, Appl1
7	1307	31.5	508	1	US-08-021-601-10 Sequence 10, Appl1
8	1307	31.5	508	1	US-08-082-849B-10 Sequence 10, Appl1
9	1307	31.5	508	5	PCT-US94-01624-10 Sequence 10, Appl1
10	1306	31.5	456	1	US-08-021-601-6 Sequence 6, Appl1
11	1306	31.5	456	1	US-08-082-849B-6 Sequence 6, Appl1
12	1306	31.5	456	5	PCT-US94-01624-6 Sequence 6, Appl1
13	504	12.2	800	6	5183745-3 Patent No. 5183745
14	228.5	5.5	186	4	US-08-973-462-8 Sequence 8, Appl1
15	208	5.0	2710	2	US-08-480-604A-6 Sequence 6, Appl1
16	208	5.0	2710	2	US-08-405-496A-6 Sequence 6, Appl1
17	208	5.0	2710	4	US-08-915-136-6 Sequence 6, Appl1
18	195	4.7	1588	5	PCT-US83-07261-11 Sequence 11, Appl1
19	195	4.7	1588	5	PCT-US93-07261-16 Sequence 16, Appl1
20	194.5	4.7	3248	1	US-08-353-700-1 Sequence 1, Appl1
21	194.5	4.7	3248	5	PCT-US95-16216-1 Sequence 1, Appl1
22	191.5	4.6	1164	4	US-08-923-992A-2 Sequence 2, Appl1
23	189	4.6	663	4	US-09-196-293-5 Sequence 5, Appl1
24	189	4.6	663	4	US-08-209-603E-5 Sequence 5, Appl1
25	188	4.5	576	4	US-09-104-324B-4 Sequence 4, Appl1
26	187	4.5	663	4	US-08-235-836C-78 Sequence 78, Appl1
27	186.5	4.5	990	2	US-08-645-193B-15 Sequence 15, Appl1

ALIGNMENTS

28	186.5	4.5	1098	4	US-08-923-992A-8	Sequence 8, Appl1
29	186.5	4.5	1164	4	US-08-923-992A-10	Sequence 10, Appl1
30	186	4.5	1285	4	US-09-308-375-2	Sequence 2, Appl1
31	184.5	4.5	1368	2	US-08-685-576-4	Sequence 4, Appl1
32	182.5	4.4	990	2	US-08-392-625-20	Sequence 20, Appl1
33	182.5	4.4	990	2	US-08-466-961A-20	Sequence 20, Appl1
34	182	4.4	2482	1	US-08-328-254-6	Sequence 6, Appl1
35	181	4.4	663	4	US-08-235-836C-70	Sequence 70, Appl1
36	181	4.4	1312	2	US-08-592-126-148	Sequence 148, App
37	181	4.4	1312	2	US-08-687-080-51	Sequence 51, Appl1
38	180.5	4.4	1128	4	US-08-923-992A-6	Sequence 6, Appl1
39	180	4.3	693	4	US-08-235-836C-72	Sequence 72, Appl1
40	179	4.3	1104	4	US-08-923-992A-4	Sequence 4, Appl1
41	176.5	4.3	1388	2	US-08-685-576-1	Sequence 1, Appl1
42	172.5	4.2	1073	4	US-09-541-782-6	Sequence 6, Appl1
43	171	4.1	872	1	US-08-766-014-2	Sequence 2, Appl1
44	168.5	4.1	700	4	US-08-235-836C-74	Sequence 74, Appl1
45	168.5	4.1	1038	4	US-09-541-782-4	Sequence 4, Appl1

```

RESULT 1
US-08-021-601-2
; Sequence 2, Application US/08021601
; Patent No. 5591631
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Kilmpel, Kurt R.
; APPLICANT: Nichols, Peter J.
; APPLICANT: Atora, Naveen
; APPLICANT: Singh, Yogendra
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,601
; FILING DATE: 19930212
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratl, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 776 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-021-601-2
Query Match 96.1%; Score 3983; DB 1; Length 776;
Best Local Similarity 100.0%; Pred. No. 66-239;
Matches 776; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 34 AGGIGDVGMHVKEKKKKDKRDEERNKTOEHLKIKHIVKIEVKOEAVKKKAEE 93

```

```

Db 1 AGGHGDVGMHVKKEKKNDEKRNKDEERNKTOEHLKEIMKHIVYIEVKGGEAVYKEAAE 60
QY KLEKVPSPDVLVEMYKAIGKIYIVGDDITKHSLEALSSEDKKKIKIDYKQDALLHEHYV 153
Db 61 KLEKVPSPDVLVEMYKAIGKIYIVGDDITKHSLEALSSEDKKKIKIDYKQDALLHEHYV 120
QY AKEGEYEVLIQSSSDYVNTKALNVYIEGKILSRDILSKINOPYKFLDVLNTIKNA 213
Db 121 AKEGEYEVLIQSSSDYVNTKALNVYIEGKILSRDILSKINOPYKFLDVLNTIKNA 180
QY SDSOGDILLFTNOLKEHPTDSVEFLLEQNSNEVQEVFAKARAYYTEPOHRYDLYAPEA 273
Db 181 SDSOGDILLFTNOLKEHPTDSVEFLLEQNSNEVQEVFAKARAYYTEPOHRYDLYAPEA 240
QY FNYMDFNEOEINLSLEELKDQRMLSRYEKMEKIKOHQHSNDSISEGRGLKKLQIPI 333
Db 241 FNYMDFNEOEINLSLEELKDQRMLSRYEKMEKIKOHQHSNDSISEGRGLKKLQIPI 300
QY EPKDDIHSLSQEEKELKKRIQIDSDPLSTEEKEFLKKLQIDIRDSLSEEEKELNRI 393
Db 301 EPKDDIHSLSQEEKELKKRIQIDSDPLSTEEKEFLKKLQIDIRDSLSEEEKELNRI 360
QY QVDSNPULSEKEKEFLKKLQIDOPYDINQRLQDTGGLIDSPSINLDVRYKQYKRDIONID 453
Db 361 QVDSNPULSEKEKEFLKKLQIDOPYDINQRLQDTGGLIDSPSINLDVRYKQYKRDIONID 420
QY ALLHOSISSTLYNKIYLENNMINNL/TATLGADLVSDTNFKINGINFEKKNKYSIS 513
Db 421 ALLHOSISSTLYNKIYLENNMINNL/TATLGADLVSDTNFKINGINFEKKNKYSIS 480
QY SNYMTVDINERPALDNERLKWRIQSPTRAGYLENGKLLIQRNIGLEIKDVOIIOKSEK 573
Db 481 SNYMTVDINERPALDNERLKWRIQSPTRAGYLENGKLLIQRNIGLEIKDVOIIOKSEK 540
QY EYIRIDAVVPKSKIDTQIOEAQLINQEWNKALGLPKYTKLITFVNHRRVASNIVESAY 633
Db 541 EYIRIDAVVPKSKIDTQIOEAQLINQEWNKALGLPKYTKLITFVNHRRVASNIVESAY 600
QY LILNEMKNNIQSDLIKKTNTNYLDGNGRFVFTDITLPLNIAEQYTHODEIYEVHSGKLYV 693
Db 601 LILNEMKNNIQSDLIKKTNTNYLDGNGRFVFTDITLPLNIAEQYTHODEIYEVHSGKLYV 660
QY PESRSILHSGSKGYELANDSEGFTHERGHAVDYAGYLLDKNOSDLYTNSKFFIDIRKE 753
Db 661 PESRSILHSGSKGYELANDSEGFTHERGHAVDYAGYLLDKNOSDLYTNSKFFIDIRKE 720
QY BGSNLTSTGRTEAEFFPAFAFRMLHSTDAERLKYOKNAPKPEOFLINDQIKIINS 809
Db 721 BGSNLTSTGRTEAEFFPAFAFRMLHSTDAERLKYOKNAPKPEOFLINDQIKIINS 776

RESULT 2
US-08-082-849B-2
; Sequence 2, Application US/08082849B
; Patent No. 5677274
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Arota, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/082,849B
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/021,601
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-161-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 776 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-082-849B-2

Query Match 96.1%; Score 3983; DB 1; Length 776;
Best Local Similarity 100.0%; Pred. No. 6e-239;
Matches 776; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGGHGDVGMHVKKEKKNDEKRNKDEERNKTOEHLKEIMKHIVYIEVKGGEAVYKEAAE 93
Db 1 AGGHGDVGMHVKKEKKNDEKRNKDEERNKTOEHLKEIMKHIVYIEVKGGEAVYKEAAE 60
QY KLEKVPSPDVLVEMYKAIGKIYIVGDDITKHSLEALSSEDKKKIKIDYKQDALLHEHYV 153
Db 61 KLEKVPSPDVLVEMYKAIGKIYIVGDDITKHSLEALSSEDKKKIKIDYKQDALLHEHYV 120
QY AKEGEYEVLIQSSSDYVNTKALNVYIEGKILSRDILSKINOPYKFLDVLNTIKNA 213
Db 121 AKEGEYEVLIQSSSDYVNTKALNVYIEGKILSRDILSKINOPYKFLDVLNTIKNA 180
QY SDSOGDILLFTNOLKEHPTDSVEFLLEQNSNEVQEVFAKARAYYTEPOHRYDLYAPEA 273
Db 181 SDSOGDILLFTNOLKEHPTDSVEFLLEQNSNEVQEVFAKARAYYTEPOHRYDLYAPEA 240
QY FNYMDFNEOEINLSLEELKDQRMLSRYEKMEKIKOHQHSNDSISEGRGLKKLQIPI 333
Db 241 FNYMDFNEOEINLSLEELKDQRMLSRYEKMEKIKOHQHSNDSISEGRGLKKLQIPI 300
QY EPKDDIHSLSQEEKELKKRIQIDSDPLSTEEKEFLKKLQIDIRDSLSEEEKELNRI 393
Db 301 EPKDDIHSLSQEEKELKKRIQIDSDPLSTEEKEFLKKLQIDIRDSLSEEEKELNRI 360
QY QVDSNPULSEKEKEFLKKLQIDOPYDINQRLQDTGGLIDSPSINLDVRYKQYKRDIONID 453
Db 361 QVDSNPULSEKEKEFLKKLQIDOPYDINQRLQDTGGLIDSPSINLDVRYKQYKRDIONID 420
QY ALLHOSISSTLYNKIYLENNMINNL/TATLGADLVSDTNFKINGINFEKKNKYSIS 513
Db 421 ALLHOSISSTLYNKIYLENNMINNL/TATLGADLVSDTNFKINGINFEKKNKYSIS 480
QY SNYMTVDINERPALDNERLKWRIQSPTRAGYLENGKLLIQRNIGLEIKDVOIIOKSEK 573
Db 541 SNYMTVDINERPALDNERLKWRIQSPTRAGYLENGKLLIQRNIGLEIKDVOIIOKSEK 540
QY EYIRIDAVVPKSKIDTQIOEAQLINQEWNKALGLPKYTKLITFVNHRRVASNIVESAY 633
Db 541 EYIRIDAVVPKSKIDTQIOEAQLINQEWNKALGLPKYTKLITFVNHRRVASNIVESAY 600
QY LILNEMKNNIQSDLIKKTNTNYLDGNGRFVFTDITLPLNIAEQYTHODEIYEVHSGKLYV 693
Db 634 LILNEMKNNIQSDLIKKTNTNYLDGNGRFVFTDITLPLNIAEQYTHODEIYEVHSGKLYV 693
```

Db 601 LILNEMKNNIQSDLIKRYNTNLYVDGNGRFVFTDITLPLNIAEQYTHODEIYEQVHSGLYV 660
QY 694 PESRSTLLHGSPKGYELRNDSEGFIEFGHAYVDVAGYLLDKNOSDLYNNSKKFIDIFE 753
Db 661 PESRSTLLHGSPKGYELRNDSEGFIEFGHAYVDVAGYLLDKNOSDLYNNSKKFIDIFE 720
QY 754 EGSNLTSGRTNEAEFFAEFRMLHSTDAERLKYOKNAPKTFPOFINOIKFIINS 809
Db 721 EGSNLTSGRTNEAEFFAEFRMLHSTDAERLKYOKNAPKTFPOFINOIKFIINS 776

RESULT 3
PCT-US94-01624-2

: Sequence 2, Application PC/TUS9401624

GENERAL INFORMATION:

APPLICANT: Leppla, Stephen H.
APPLICANT: Kilmpel, Kurt R.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE and CREM
STREET: Stewart Street Tower, 20th floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 776 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-2

Query Match 96.1%; Score 3983; DB 5; Length 776;

Best Local Similarity 100.0%; Pred. No. 6e-239;
Matches 776; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGCHDVGCMHYKEREKKNKDEKRNKTOEHLKETIMKHIYKIEVGEAEVKKAE 93
Db 1 AGCHDVGCMHYKEREKKNKDEKRNKTOEHLKETIMKHIYKIEVGEAEVKKAE 60
QY 94 KLEKVPDVLMEYKAIGKIYVDGDTKHISLEALSEDKKKIKDIYGRDALHHEVY 153
Db 61 KLEKVPDVLMEYKAIGKIYVDGDTKHISLEALSEDKKKIKDIYGRDALHHEVY 120
QY 154 AKEGEYPLVYQSSRDYVNTPEKALNVYIEIKIISRLSLKINOPYQGFLLVNTIKNA 213
Db 121 AKEGEYPLVYQSSRDYVNTPEKALNVYIEIKIISRLSLKINOPYQGFLLVNTIKNA 180
QY 214 SDSGODLLFTNLQKHPDFSVFLEQNSNEQVEFAFAFAYIEPQRHVDLQLYAPBA 273

Db 181 SDSGODLLFTNLQKHPDFSVFLEQNSNEQVEFAFAFAYIEPQRHVDLQLYAPBA 240
QY 274 FNYMDFNEOEINLSLEELKDQRMLSRYEKWEKIKOHYHMSDLSSEBGRGLKKLOIPI 333
Db 241 FNYMDFNEOEINLSLEELKDQRMLSRYEKWEKIKOHYHMSDLSSEBGRGLKKLOIPI 300
QY 334 EPKPDIIHLSOEKELKRIODSSDFLSTEEKFLLKLOIDIRDSISEEKKELNFI 393
Db 301 EPKPDIIHLSOEKELKRIODSSDFLSTEEKFLLKLOIDIRDSISEEKKELNFI 360
QY 394 QVDSNPLSEKEKEFKLKLIDQPYDINQRLQDTGGLIDSPSINDVRYKQYKRDIONID 453
Db 361 QVDSNPLSEKEKEFKLKLIDQPYDINQRLQDTGGLIDSPSINDVRYKQYKRDIONID 420
QY 454 ALIHQISGTLVYKILYENMNINNLJATLGADLVSTONTKINRGIENEFKKNFYSIS 513
Db 421 ALIHQISGTLVYKILYENMNINNLJATLGADLVSTONTKINRGIENEFKKNFYSIS 480
QY 514 SNMIVDINERPALDNERLKWRTQLSPPDRAGLENGKLLILQNIISLEIKVOIIOSEK 573
Db 481 SNMIVDINERPALDNERLKWRTQLSPPDRAGLENGKLLILQNIISLEIKVOIIOSEK 540
QY 574 EYRIDAKVPSKIDTKIOEAOININQEMNKALGPKYTKLTFVYHNRYSNIVESAY 633
Db 541 EYRIDAKVPSKIDTKIOEAOININQEMNKALGPKYTKLTFVYHNRYSNIVESAY 600
QY 634 LILNEMKNNIQSDLIKRYNTNLYVDGNGRFVFTDITLPLNIAEQYTHODEIYEQVHSGLYV 693
Db 601 LILNEMKNNIQSDLIKRYNTNLYVDGNGRFVFTDITLPLNIAEQYTHODEIYEQVHSGLYV 660
QY 694 PESRSTLLHGSPKGYELRNDSEGFIEFGHAYVDVAGYLLDKNOSDLYNNSKKFIDIFE 753
Db 661 PESRSTLLHGSPKGYELRNDSEGFIEFGHAYVDVAGYLLDKNOSDLYNNSKKFIDIFE 720
QY 754 EGSNLTSGRTNEAEFFAEFRMLHSTDAERLKYOKNAPKTFPOFINOIKFIINS 809
Db 721 EGSNLTSGRTNEAEFFAEFRMLHSTDAERLKYOKNAPKTFPOFINOIKFIINS 776

RESULT 4

: US-08-021-601-8
: Sequence 8, Application US/08021601
: Patent No. 5591631

GENERAL INFORMATION:

APPLICANT: Leppla, Stephen H.
APPLICANT: Kilmpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057

TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-8

Query Match 31.6%; Score 1309; DB 1; Length 472;
Best Local Similarity 98.8%; Pred. No. 9,4e-74;
Matches 255; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 34 AGGHDVGMHVKKEKKDKDEERNKTOEHLKIMKHIVKIEVKEEAVKKEAAE 93
|
DB 4 AGGHDVGMHVKKEKKDKDEERNKTOEHLKIMKHIVKIEVKEEAVKKEAAE 63
QY 94 KLEKVPDVLVEMKKAIGKIYVDGDTKHSLEALSEDEKKIKIDYGDALLHEHYV 153
|
DB 64 KLEKVPDVLVEMKKAIGKIYVDGDTKHSLEALSEDEKKIKIDYGDALLHEHYV 123
QY 154 AKEGEPVLVIOSEEDYVENTERKALNYYEIGKILSRDILSKINOPYOKFLDVLNTTKNA 213
|
DB 124 AKEGEPVLVIOSEEDYVENTERKALNYYEIGKILSRDILSKINOPYOKFLDVLNTTKNA 183
QY 214 SDSGDDLLFTNOLKEHPTFSVEFLQNSNEVOEFAKAFAYITEPOHNDVLOLYAPEA 273
|
DB 184 SDSGDDLLFTNOLKEHPTFSVEFLQNSNEVOEFAKAFAYITEPOHNDVLOLYAPEA 243
QY 274 FNYMDKFNQEQEINLSLEE 291
|
DB 244 FNYMDKFNQEQEINLTRAEE 261

RESULT 5
US-08-082-849B-8
Sequence 8, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Kimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-8

Query Match 31.6%; Score 1309; DB 1; Length 472;
Best Local Similarity 98.8%; Pred. No. 9,4e-74;
Matches 255; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 34 AGGHDVGMHVKKEKKDKDEERNKTOEHLKIMKHIVKIEVKEEAVKKEAAE 93
|
DB 4 AGGHDVGMHVKKEKKDKDEERNKTOEHLKIMKHIVKIEVKEEAVKKEAAE 63
QY 94 KLEKVPDVLVEMKKAIGKIYVDGDTKHSLEALSEDEKKIKIDYGDALLHEHYV 153
|
DB 64 KLEKVPDVLVEMKKAIGKIYVDGDTKHSLEALSEDEKKIKIDYGDALLHEHYV 123
QY 154 AKEGEPVLVIOSEEDYVENTERKALNYYEIGKILSRDILSKINOPYOKFLDVLNTTKNA 213
|
DB 124 AKEGEPVLVIOSEEDYVENTERKALNYYEIGKILSRDILSKINOPYOKFLDVLNTTKNA 183
QY 214 SDSGDDLLFTNOLKEHPTFSVEFLQNSNEVOEFAKAFAYITEPOHNDVLOLYAPEA 273
|
DB 184 SDSGDDLLFTNOLKEHPTFSVEFLQNSNEVOEFAKAFAYITEPOHNDVLOLYAPEA 243
QY 274 FNYMDKFNQEQEINLSLEE 291
|
DB 244 FNYMDKFNQEQEINLTRAEE 261

RESULT 6
PCT-US94-01624-8
Sequence 8, Application PC/US9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Kimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: Steuart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-8

Query Match 31.6%; Score 1309; DB 5; Length 472;
Best Local Similarity 98.8%; Pred. No. 9,4e-74;
Matches 255; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 34 AGHGQVGMHVKKEKNKDEERNKTOEHLKEIMKHIVKTEVGEAVKKEAAE 93
DB 4 AGHGQVGMHVKKEKNKDEERNKTOEHLKEIMKHIVKTEVGEAVKKEAAE 63
OY 94 KLEKVPDYLEMYKAIGGIYIVDGTIKHISLEALSEDKKIKIDYKGDALLHEHYV 153
DB 64 KLEKVPDYLEMYKAIGGIYIVDGTIKHISLEALSEDKKIKIDYKGDALLHEHYV 123
OY 154 AKEGYEPVLIQSSSEYVENTEKALNYYEIGKILSRDLSKINOPYOKFLDVLNTIKNA 213
DB 124 AKEGYEPVLIQSSSEYVENTEKALNYYEIGKILSRDLSKINOPYOKFLDVLNTIKNA 183
OY 214 SPSDQDGLFTNQLKEHPTDFSVFLEQNSNEVOEYFAKAFAYIIEPQHRDVLQYAPEA 273
DB 184 SPSDQDGLFTNQLKEHPTDFSVFLEQNSNEVOEYFAKAFAYIIEPQHRDVLQYAPEA 243
OY 274 FNYMDKFNEQEIINLSLE 291
DB 244 FNYMDKFNEQEIINLTRA 261

RESULT 7
US-08-021-601-10
Sequence 10, Application US/08021601
Patent No. 5591631

GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414, 057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-9880
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 508 amino acids

TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-10

Query Match 31.5%; Score 1307; DB 1; Length 508;
Best Local Similarity 99.6%; Pred. No. 1,4e-73;
Matches 254; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 34 AGHGQVGMHVKKEKNKDEERNKTOEHLKEIMKHIVKTEVGEAVKKEAAE 93
DB 1 AGHGQVGMHVKKEKNKDEERNKTOEHLKEIMKHIVKTEVGEAVKKEAAE 60
OY 94 KLEKVPDYLEMYKAIGGIYIVDGTIKHISLEALSEDKKIKIDYKGDALLHEHYV 153
DB 61 KLEKVPDYLEMYKAIGGIYIVDGTIKHISLEALSEDKKIKIDYKGDALLHEHYV 120
OY 154 AKEGYEPVLIQSSSEYVENTEKALNYYEIGKILSRDLSKINOPYOKFLDVLNTIKNA 213
DB 121 AKEGYEPVLIQSSSEYVENTEKALNYYEIGKILSRDLSKINOPYOKFLDVLNTIKNA 180
OY 214 SPSDQDGLFTNQLKEHPTDFSVFLEQNSNEVOEYFAKAFAYIIEPQHRDVLQYAPEA 273
DB 181 SPSDQDGLFTNQLKEHPTDFSVFLEQNSNEVOEYFAKAFAYIIEPQHRDVLQYAPEA 240
OY 274 FNYMDKFNEQEIINLS 288
DB 241 FNYMDKFNEQEIINLT 255

RESULT 8
US-08-082-849B-10
Sequence 10, Application US/08082849B
Patent No. 5677274

GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 508 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-10

Query Match 31.5%; Score 1307; DB 1; Length 508;
Best Local Similarity 99.6%; Pred. No. 1,4e-73;
Matches 254; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGGHQDVGMHVKKEKKNDEKRRKDEERNKTOEHLKEIMKHIVKEVGEAVKKEAAE 93
|||||
DB 1 AGGHQDVGMHVKKEKKNDEKRRKDEERNKTOEHLKEIMKHIVKEVGEAVKKEAAE 60
QY 94 KLEKVPSPDLVEMKKAIGKIYVGDITKHSLEALSEDKKKIKIDYGRKALLHEHYV 153
|||||
DB 61 KLEKVPSPDLVEMKKAIGKIYVGDITKHSLEALSEDKKKIKIDYGRKALLHEHYV 120
QY 154 AKEGEPLVLIQSSSEDIYENTKALNYYEIGKILSRILSKINOPYOKFLDVLNTIKNA 213
|||||
DB 121 AKEGEPLVLIQSSSEDIYENTKALNYYEIGKILSRILSKINOPYOKFLDVLNTIKNA 180
QY 214 SDSGQDLFTNOLKEHPTDSVEFLQNSNEVOEVFAKAFAYIIEPQHRDVLQYAPEA 273
|||||
DB 181 SDSGQDLFTNOLKEHPTDSVEFLQNSNEVOEVFAKAFAYIIEPQHRDVLQYAPEA 240
QY 274 FNYMDKFNQOEINLS 288
|||||
DB 241 FNYMDKFNQOEINLT 255

RESULT 9
PCT-US94-01624-10
Sequence 10, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND KHOURI and CREW
STREET: Stewart Street Tower, 20th floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 508 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

PCT-US94-01624-10

Query Match 31.5%; Score 1307; DB 5; Length 508;
Best Local Similarity 99.6%; Pred. No. 1,4e-73;
Matches 254; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGGHQDVGMHVKKEKKNDEKRRKDEERNKTOEHLKEIMKHIVKEVGEAVKKEAAE 93
|||||
DB 1 AGGHQDVGMHVKKEKKNDEKRRKDEERNKTOEHLKEIMKHIVKEVGEAVKKEAAE 60
QY 94 KLEKVPSPDLVEMKKAIGKIYVGDITKHSLEALSEDKKKIKIDYGRKALLHEHYV 153
|||||
DB 61 KLEKVPSPDLVEMKKAIGKIYVGDITKHSLEALSEDKKKIKIDYGRKALLHEHYV 120
QY 154 AKEGEPLVLIQSSSEDIYENTKALNYYEIGKILSRILSKINOPYOKFLDVLNTIKNA 213
|||||
DB 121 AKEGEPLVLIQSSSEDIYENTKALNYYEIGKILSRILSKINOPYOKFLDVLNTIKNA 180
QY 214 SDSGQDLFTNOLKEHPTDSVEFLQNSNEVOEVFAKAFAYIIEPQHRDVLQYAPEA 273
|||||
DB 181 SDSGQDLFTNOLKEHPTDSVEFLQNSNEVOEVFAKAFAYIIEPQHRDVLQYAPEA 240
QY 274 FNYMDKFNQOEINLS 288
|||||
DB 241 FNYMDKFNQOEINLT 255

RESULT 10
US-08-021-601-6
US-08-021-601-6
Sequence 6, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414,057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-6

Query Match 31.5%; Score 1306; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 1,4e-73;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGGHDVGAMHVKKEKKNKDENRKDEERKKTQEBHLKEIMKHIVIEVKGEEAVKKEAAE 93
DB 1 AGGHGDVGMHVKKEKKNKDENRKDEERKKTQEBHLKEIMKHIVIEVKGEEAVKKEAAE 60
QY 94 KLEEVPSVLEMYKAIGKITYVDGDTKHTSLALSDDKKIKIDYKGDALLHEHHY 153
DB 61 KLEEVPSVLEMYKAIGKITYVDGDTKHTSLALSDDKKIKIDYKGDALLHEHHY 120
QY 154 AKEGEYPLVIOSSSEDYVENTEKALNVYIEIGKILSRDLSKINOPYKFLDVLNTIKNA 213
DB 121 AKEGEYPLVIOSSSEDYVENTEKALNVYIEIGKILSRDLSKINOPYKFLDVLNTIKNA 180
QY 214 SDSGODLLFTNQLKEHPTDFSVFLEONSNEVOEFAKAFAYYIEPOHRDVLQLYAPEA 273
DB 181 SDSGODLLFTNQLKEHPTDFSVFLEONSNEVOEFAKAFAYYIEPOHRDVLQLYAPEA 240
QY 274 FNYMDKFNEOEINL 287
DB 241 FNYMDKFNEOEINL 254

RESULT 11

US-08-082-849B-6
Sequence 6, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arota, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-6

Query Match 31.5%; Score 1306; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 1,4e-73;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGGHDVGAMHVKKEKKNKDENRKDEERKKTQEBHLKEIMKHIVIEVKGEEAVKKEAAE 93
DB 1 AGGHGDVGMHVKKEKKNKDENRKDEERKKTQEBHLKEIMKHIVIEVKGEEAVKKEAAE 60
QY 94 KLEEVPSVLEMYKAIGKITYVDGDTKHTSLALSDDKKIKIDYKGDALLHEHHY 153
DB 61 KLEEVPSVLEMYKAIGKITYVDGDTKHTSLALSDDKKIKIDYKGDALLHEHHY 120
QY 154 AKEGEYPLVIOSSSEDYVENTEKALNVYIEIGKILSRDLSKINOPYKFLDVLNTIKNA 213
DB 121 AKEGEYPLVIOSSSEDYVENTEKALNVYIEIGKILSRDLSKINOPYKFLDVLNTIKNA 180
QY 214 SDSGODLLFTNQLKEHPTDFSVFLEONSNEVOEFAKAFAYYIEPOHRDVLQLYAPEA 273
DB 181 SDSGODLLFTNQLKEHPTDFSVFLEONSNEVOEFAKAFAYYIEPOHRDVLQLYAPEA 240
QY 274 FNYMDKFNEOEINL 287
DB 241 FNYMDKFNEOEINL 254

RESULT 12

PCT-US94-01624-6
Sequence 6, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arota, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: Steuart Street Tower, 20th Floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-6

Query Match 31.5%; Score 1306; DB 5; Length 456;
Best Local Similarity 100.0%; Pred. No. 1,4e-73;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
Db 1063 LNKLENISSGVEQVTEHVEQNVYVDVPAKMQDFGLINEAGLKMENFNLVDVK 1122
Qy 212 NMSDSGOLLTNNQLEKPTDSVEFLQNSNEVOEFAKAFAYIEQHRDVLQVAP 271
Db 1123 SES-----DVITVEEIKDER-----VOKEVEKETVSIIEEMEEIYDVLEE 1163
Qy 272 EAFNMYDKF-----NEDEINLSLEELKQORLSREKEKEKIK 308
Db 1164 EREDLTKMIDAVESEIETSSDSKETESIKQEKQVSLVEEVDNDDESVEKLEK 1223
Qy 309 OHYOHV-----SDLSSEGRGL-----LKKQIPIE-----PKKD 338
Db 1224 NMEELMKDAVEINDITSLEIETQELNEVADLIKDMEKLEKALBEDSKEIIDA 1283
Qy 339 DIHLSIOBEKEL-----LKRQIDS-----SDFLSTEE-----KEF----- 370
Db 1284 DYLEKVEIEEHDTITLDEVLEKQVEEDKIEKVSCLKDEEDILKEVEKEIESEILE 1343
Qy 371 -----LKKLQIDIRDSLS-----EBEKELLNRIQVDSNPSEKEKFLKTKLQI 416
Db 1344 DYKELKTITDLEEKKEIEKHFEKEFEBAEIKDLEADILKEVSSLEVEEKKLE--- 1400
Qy 417 QPYDINQRLQD-----TGGLIDSPSINLDVRKQYKRDQIONIDA 454
Db 1401 EYHELKEEVEHIIISGDANIKGLEEDLEEVDDLKGSILMLKGMELGMDKESLEDVTT 1460
Qy 455 LHOISGSLYKIKYILEMMNINLTATGADLVSTDMTKINRG-----IFNEKKNF 508
Db 1461 KGEREVESLAK-----DVLSSALGMDDEQMKTRKKAQRPKLEEVLLAKEVEKEP 1508
Qy 509 KXSISNNYIVDINERPALD-----NEBRLKMRIO--LSPDTRAGYLENGKLLIQRI 558
Db 1509 KKITKKKKKRPDIKQEPDEIVEEMKDEIDIEEDVEEDIEEDKVEDIDEDIDEDI 1568
Qy 559 GLEIKD--VOIKOSEKEVIRIDAKVVPKSIDTKIOEQLNINQEMKALGPKYTKLI 616
Db 1569 G-EDDEVIDLIVQKEKRIEKYKAK---KKLEKKEVEGVSGLKKHVDYEM---KVVQKI 1621
Qy 617 TENVHNRVASNIVESAYLLENKKNINOSDLIKKVTNYL 655
Db 1622 DKEV-DKEVSKALESKNDVTNVLKON--ODPFSKVKNFV 1657

RESULT 15
; Sequence 6, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-604A-6
```

```
Query Match 5.0%; Score 208; DB 1; Length 2710;
Best Local Similarity 20.4%; Pred. No. 0.00011;
Matches 196; Conservative 139; Mismatches 277; Indels 348; Gaps 51;

Qy 48 EKKKDKNRKDEERKNTQOEHLKET-MKHIVKIEVKG--EEVYKKEAEKLEK----V 99
Db 212 EYNRDETVLESYRTNS-----LRKINSNHGIDRANSFLTEQELNIYSEELNRGNLAA 266
Qy 100 PSDVLEM--YKATGCKIYI-----VDDGITKHS-----LEALSDKKK 136
Db 267 ASDIVRLALAKNFG--VYLDVDMLPQIHSDLEFTTISRPSIGLDRRDMKLEKAIKRYK 325
Qy 137 IKDIYKDALLHEHYVANEQEPV-----LVYQSEEDYVENTERALN----- 179
Db 326 INN-----YTSNFPKLDQQLKDNFKLIESEKSEKSEIFSKLENLNVSDLEI 372
Qy 180 -VYIEIGKILSRDILK-----INOPYQKFLDVLNTIKNASDSCODLLFTNOLK- 228
Db 373 KIAFALGSVINQALISKOSYLTNLVIEQVKNYQFLNQHNLNPAIESDNN--FTDTTKI 429
Qy 229 EHPDTSVERLEQNSNEVOEFAKAFAYIE-----POHRDVLQVAPFAF--NYMDKPN 282
Db 430 FHDLSFNATFENS-----MFLTKTAPLYQVCFMPEANSTJLSGGAASAYYDFINL 483
Qy 283 OEINL-----SLEELKQORLSRYE-----KWEKIKQHYQMSDS 317
Db 484 QENTIKETLKASDLIEFKFPENNLSOLTEQELNSLMSPOASAKYQFQEKVRYDT--GGS 541
Qy 318 LSEEG-----RGLKTKLQIP-----IEPKKDI-----IHS 343
Db 542 LSEDNQVDNPKNTALDKNYLNNKIPSNVVEAGSKNYHYIIQLOGDDISYEATCNLFS 601
Qy 344 LSOEKEKLLKRIQIDS--SDFLSTEEKFKLKKQIDIRDSLEKEKELN----- 391
Db 602 KNPKNSTIIOIRNNNESAKSFUSDGDESTLELKKYRIPKLNKKKVKVYKTFIGHGDEFN 661
Qy 392 -----RIQVDSNPSEKEKFLKTKLQIDP-----YDINQRLQDQGLI 432
Db 662 TSEFARLSVDS--LSNETSSFLDTIKLIDISPRKNVENVNLGCMFSDYDNVEETVYCKLL 718
```

```

0Y 433 DS-----PSINIDV-----KROKRDION-----IDALL-----OSIGSLYKI 468
    | : : : : : | : : : : : | : : : : : | : : : : : |
Db 719 LSIIDKTTSTLPDVKNKSSITIGANOIVRINSCKRELLASHGCKMIKEALMSLSCKE 778
    | : : : : : | : : : : : | : : : : : | : : : : : |
0Y 469 YLYENMNINLLTATLGADLVSDTDNRKINRGIFNEFKKNFKXSISNW-IVDINERPA 526
    | : : : : : | : : : : : | : : : : : | : : : : : |
Db 779 YIF-----PDSIDN-----KIKAKSKNIPGLASISE--- 804
    | : : : : : | : : : : : | : : : : : | : : : : : |
0Y 527 LDNRRLKWRIOLOSPDTAGLNGCKLLQNRIGLEFKDVOITIKOSEKEYIRID---AKV 582
    | : : : : : | : : : : : | : : : : : | : : : : : |
Db 805 -DITFLLLDASVSPTFK-ILNNLKINIESIS- -DIYEEKLEPVKN 848
    | : : : : : | : : : : : | : : : : : | : : : : : |
0Y 583 VPKSIDTKIOEAOI---NINOENMKALGLPKYTKLTFPNVNRNRYASNTEASAYLLNEMK 640
    | : : : : : | : : : : : | : : : : : | : : : : : |
Db 849 ITHSIDDLIDELFULENVSD- -LYELKLL- -NNIDEKYLISFE- 889
    | : : : : : | : : : : : | : : : : : | : : : : : |
0Y 641 NNIOSDLIKKVTNLY- -DGNRGREVFDTILPNIAEQYTHODEIYBQVHSGKGLYBES 696
    | : : : : : | : : : : : | : : : : : | : : : : : |
Db 890 -DISKNNSFYSVAFIKNSGSESYVE-TEKEIFYSK- -EHUTREISF- 933
    | : : : : : | : : : : : | : : : : : | : : : : : |
0Y 697 RSILLHOPSKGVELRANDESEFIEHFGHAADDVAGYLLDKNOSDLYT-----NSKFEIDIF 751
    | : : : : : | : : : : : | : : : : : | : : : : : |
Db 934 - - - - -IKNS- - - - -IITDVGNMLNDIOIDHTSOYNTLMAAFIOSL 970
    | : : : : : | : : : : : | : : : : : | : : : : : |
0Y 752 KEFGSN- - -LTSYGRTEAEFFAEFRMHST- -DHAERLKYOXNAPPTQFINDOIKFI 806
    | : : : : : | : : : : : | : : : : : | : : : : : |
Db 971 IDYSKNDVLDNDLSTSKVQVLYAOLESTGINTIYDSIOVLVNLISNA- - -VNDTTIYVL 1024
    | : : : : : | : : : : : | : : : : : | : : : : : |

```

Search completed: December 2, 2001, 13:49:10
Job time: 190 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2001, 13:46:00 ; Search time 112.1 Seconds
(without alignments)
534.570 Million cell updates/sec

Title: US-09-747-521-2

Perfect score: 4145
Sequence: 1 MNKKEFKIVSMCLVTAI.....KNAPKTFQINDQIFILNS 809

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101:*

- 1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT:*
- 22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4145	100.0	809	22	AA1980
2	3983	96.1	776	15	AA1981
3	1322	31.9	485	22	AA1982
4	1309	31.6	472	15	AA1983
5	1307	31.5	508	15	AA1984
6	1306	31.5	456	15	AA1985
7	488	11.8	800	11	AA1986
8	264.5	6.4	1979	21	AA1987
9	235.5	5.7	2663	22	AA1988
10	235.5	5.7	2688	22	AA1989
11	228.5	5.5	1786	18	AA1990

12	224.5	5.4	1558	21	AA18324	Plasmodium falciparum
13	218	5.3	980	21	AA18294	Plasmodium falciparum
14	215	5.2	1145	22	AA18216	S. epidermidis ope
15	213	5.1	1516	21	AA18195	Plasmodium falciparum
16	211	5.1	1254	18	AA18172	Merizote apical-en
17	211	5.1	1254	18	AA18172	Merizote apical-en
18	209.5	5.1	1780	22	AA18172	Human polypeptide
19	208	5.0	2710	17	AA18172	Human polypeptide
20	208	5.0	2710	17	AA18172	C. difficile toxin
21	204	4.9	1866	19	AA18172	Plasmodium falciparum
22	203	4.9	2485	21	AA18172	Plasmodium falciparum
23	202.5	4.9	1788	22	AA18172	Human polypeptide
24	198.5	4.8	789	22	AA18172	Human polypeptide
25	198.5	4.8	1427	12	AA18172	Human 160k mediat
26	198	4.8	497	20	AA18172	B. burgdorferi ant
27	196	4.7	481	20	AA18172	B. burgdorferi ant
28	195.5	4.7	630	18	AA18172	Mouse RHAMM protel
29	195.5	4.7	1392	20	AA18172	Restin protein seq
30	195	4.7	1588	15	AA18172	Malarial PFEPM3 ep
31	195	4.7	1663	15	AA18172	Plasmodium falciparum
32	194.5	4.7	3248	17	AA18172	Kinetochore protel
33	194	4.7	850	19	AA18172	A. thermophilum th
34	194	4.7	5024	22	AA18172	S. epidermidis ope
35	193	4.7	1316	22	AA18172	Maize RAD50. Zea
36	193	4.7	2206	21	AA18172	Plasmodium falciparum
37	192.5	4.6	558	21	AA18172	Plasmodium falciparum
38	191.5	4.6	1119	20	AA18172	B. burgdorferi ant
39	191.5	4.6	1164	17	AA18172	Group B streptococ
40	191.5	4.6	1164	19	AA18172	Group B streptococ
41	191.5	4.6	1164	21	AA18172	Amino acid sequenc
42	191	4.6	2440	18	AA18172	H. pylori cytoplas
43	190.5	4.6	1164	19	AA18172	Mutant C-beta prot
44	189	4.6	663	12	AA18172	B. burgdorferi stra
45	189	4.6	725	18	AA18172	Human RHAMM protel

ALIGNMENTS

RESULT 1	
AA1980	standard; Protein: 809 AA.
ID	AA1980
XX	
AC	AA1980
XX	
DT	29-AUG-2001 (first entry)
XX	
DE	Wild type B. anthracis lethal factor.
KW	Lethal factor; LF; immunogen; LF4; protective antigen; PA; DNA vaccine;
KW	humoral; cell-mediated; immune memory response.
XX	
XX	Bacillus anthracis.
XX	
FH	Key
FT	Peptide
FT	1..33
FT	/label= signal peptide
FT	/note= "Not given in the specification"
FT	34..809
FT	/label= LF
FT	42..285
FT	/label= LF4
FT	
XX	
PN	WO200145639-A2.
XX	
PD	28-JUN-2001.
XX	
PD	21-DEC-2000; 2000WO-US34912.
XX	
PR	22-DEC-1999; 99US-0171459.
XX	
PA	(OHIS) UNIV OHIO STATE RES FOUND.
PA	(GALL/) GALLOWAY D R.

PA (MATE/) MATECZUN A J.
XX
XX Galloway DR, Mateczun AJ;
XX
DR WPI: 2001-408540/43.
XX N-PSDB: AAC86015.
XX
PT Protecting animal against lethal infection with Bacillus anthracis, by
PT administering wild type or mutated form of Bacillus anthracis lethal
PT factor protein or its fragment or a nucleic acid encoding the mutated
PT protein -
PS
PS Claim 3; Fig 1; 33pp: English.
XX
XX This sequence represents the B. anthracis lethal factor (LF). An
CC immunogenic fragment of LF, LF4, can be used to produce an immune
CC response which protects an animal against lethal infection with
CC Bacillus anthracis. DNA encoding the B. anthracis LF can be used
CC in conjunction with DNA encoding the protective antigen (PA) in a
CC DNA vaccine. Using a DNA vaccine which encodes the mutated LF protein
CC or fragment alone or in combination with a DNA encoding the PA protein
CC or its fragment, both components (humoral and cell-mediated) of the
CC immune system are stimulated, which results in longer term immune
CC memory response. The combined use of a mutated LF and PA gene or their
CC fragments results in a higher level of immune response, as judged by
CC overall serum antibody titers for LF and PA antigens, than the use of
CC either LF or PA genes in separate immunizations.
XX
SQ Sequence 809 AA:

Query Match 100.0%; Score 4145; DB 22; Length 809;
Best Local Similarity 100.0%; Pred. No. 2.9e-245;
Matches 809; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNIKKEFFIVISMSCLVTAITLSPVFIPLYOGAGCHGVGMHVKEKKNKDEKRDDEE 60
DB 1 mnikkeffivismsclvtaitlsgvpfiplyvgagghgvgmhvkekenkndekndee 60

QY 61 RNKTOEHLKETIMKHIVKIEVKGEEAVKKEAEKLEKVPDVLNEMYKAIIGKRIYVDCD 120
DB 61 rnktoehleketimkhivkievkggeavkkeaekllekvpdvllemnkykaiigkriyvdc 120

QY 61 rnkqtgehlkeimkhivklevkgeavkkeaekllekvpdvllemnkykaiigkriyvdc 120
DB 61 rnkqtgehlkeimkhivklevkgeavkkeaekllekvpdvllemnkykaiigkriyvdc 120

QY 121 IRKHTISLEASDCKKIKRIYKGDALHHEHYVYAKGEGYPLVYOSSEYVENTEKALNV 180
DB 121 irkhtisleasdckkikriykgdallhehyvayakgegyplvyosseyventekaln 180

QY 121 lckhstleasckkikikidallhehyvayakegyplvyosseyventekaln 180
DB 121 lckhstleasckkikikidallhehyvayakegyplvyosseyventekaln 180

QY 181 YVEIGKILSRDILSKINQYQKFLDVLTINAKNSDSGDGLFTNOLKHPHDFSVETLE 240
DB 181 yveigkilSRDILSKINQYQKFLDVLTINAKNSDSGDGLFTNOLKHPHDFSVETLE 240

QY 181 yveigkilSRDILSKINQYQKFLDVLTINAKNSDSGDGLFTNOLKHPHDFSVETLE 240
DB 181 yveigkilSRDILSKINQYQKFLDVLTINAKNSDSGDGLFTNOLKHPHDFSVETLE 240

QY 241 ONSNEVOEYFAKAFAYYIEPDRDLQLYAPEAFNTMDKFNEOEINMLSEELKQRMISR 300
DB 241 onsnevoeyfakafayyiepdrdlqlyapeafntmdkfneoeinmlseelkqrmisr 300

QY 241 qnsnevevfakafayyiepdrdlqlyapeafntmdkfneoeinmlseelkqrmisr 300
DB 241 qnsnevevfakafayyiepdrdlqlyapeafntmdkfneoeinmlseelkqrmisr 300

QY 301 YKWEKIKOHYHMSDSISEEGGILKTIQIPEPKKDDIHSISEEKEELKRIQIDSS 360
DB 301 ykwekikohyHMSDSISEEGGILKTIQIPEPKKDDIHSISEEKEELKRIQIDSS 360

QY 301 ykwekikohyHMSDSISEEGGILKTIQIPEPKKDDIHSISEEKEELKRIQIDSS 360
DB 301 ykwekikohyHMSDSISEEGGILKTIQIPEPKKDDIHSISEEKEELKRIQIDSS 360

QY 361 DELSTEKEEFLKQLQIDIRDSISEEKEELNRIQVDSNPISKEKEFLKTLQIDQPD 420
DB 361 delstekeeFLKQLQIDIRDSISEEKEELNRIQVDSNPISKEKEFLKTLQIDQPD 420

QY 361 dflsteekelfkqlqidirdsiseekellnriqvdsnpiskekeflktlqidqpd 420
DB 361 dflsteekelfkqlqidirdsiseekellnriqvdsnpiskekeflktlqidqpd 420

QY 421 INORLODTGGLDPSINIDVRKQYKRDQIONDALLHQSISGFLVKNKIYLVENMINNTLT 480
DB 421 inorlodtGGLDPSINIDVRKQYKRDQIONDALLHQSISGFLVKNKIYLVENMINNTLT 480

QY 421 lngriqldtggldpsindvrkqykrdignidallhgsigstlykilylenminnlt 480
DB 421 lngriqldtggldpsindvrkqykrdignidallhgsigstlykilylenminnlt 480

QY 481 ATLGLADVSTNTKINRGIFNEPFKNKFKYSSNMYIWDINERPLDNERLQKRIQLSP 540
DB 481 atlgladvstntkinrgifnepfknkfkYSSNMYIWDINERPLDNERLQKRIQLSP 540

QY 481 atlgldadvstntkinrgifnepfknkfkYSSNMYIWDINERPLDNERLQKRIQLSP 540
DB 481 atlgldadvstntkinrgifnepfknkfkYSSNMYIWDINERPLDNERLQKRIQLSP 540

QY 541 DTRAGTLENGKILQINILETIKDVQIIOKSEKEYIRIDAKVVPKSIDTKIOEADLNIN 600
DB 541 dtragtLENGKILQINILETIKDVQIIOKSEKEYIRIDAKVVPKSIDTKIOEADLNIN 600

*
4

DB 541 dtragtLENGKILQINILETIKDVQIIOKSEKEYIRIDAKVVPKSIDTKIQEAGLIN 600
QY 601 OEWNKALGLPKYTKLITTFVWVHNRYASNIVESAYLILNEKNMNIQSDILIKVTNYLVDGNC 660
DB 601 oewnkalglpkytklittfvwvhnryasnivesaylilneknmniqsdilikvtnylvdgnc 660

QY 661 GEWNKALGLPKYTKLITTFVWVHNRYASNIVESAYLILNEKNMNIQSDILIKVTNYLVDGNC 660
DB 661 gewnkalglpkytklittfvwvhnryasnivesaylilneknmniqsdilikvtnylvdgnc 660

QY 661 RYFTDITLTPNABOYTHODEIYEVHSGGLVPPESRSTLLHGSPSGVELRNDSECFIHE 720
DB 661 rftdiltlpnaboythodeiyevhsgglvppesrstllhgspsgvelrndsfcfihe 720

QY 721 FGHAVDVGAYLLDKNQSDLVNNSKKFIDIPKEEGSNLTSYGRTEAEFFAEAFRLMHS 780
DB 721 fghavdvgaylldknqsdlvnnskkfidipkeegsnltsygrteaeffaeafrlmhs 780

QY 781 DHAERLKVONAKPTPOFINDOIKFTINS 809
DB 781 dhaerlkvonakptpofindoiKFTINS 809

RESULT 2
AAR60178
ID AAR60178 standard; Protein; 776 AA.
XX
XX AAR60178;
AC AAR60178;
XX 03-APR-1995 (first entry)
DT
DT
XX
XX Lethal factor of Bacillus anthracis.
DE
DE Anthrax: Bacillus anthracis; fusion protein; lethal factor;
KW protective antigen; cell killing; targeting; targeting; pathogen;
KW intracellular; HIV; human immunodeficiency virus; toxin.
XX
XX Bacillus anthracis.
OS
XX
XX W09418332-A.
PN
XX
XX 18-AUG-1994.
PD
XX
XX 14-FEB-1994; 94WO-0501624.
XX
XX
XX 12-FEB-1993; 93US-0021601.
PR
XX 25-JUN-1993; 93US-0082849.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA
PI
PI Arora N, Kimpel K, Leppia SH, Nichols PJ, Singh Y;
DR WPI: 1994-279753/34.
DR N-PSDB: AA070179.
XX
XX Nucleic acid encoding anthrax toxin fusion protein - useful for
PT targeting toxin to specific cells, eg for killing tumour cells
PT or HIV-infected cells
PT
XX
PS Disclosure; Page 75-77; 124pp: English.
XX
XX The sequence encoding the lethal factor of Bacillus anthracis may be
CC used in the construction of a nucleic acid which encodes a fusion
CC protein comprising the anthrax protective antigen binding domain of
CC the native anthrax lethal factor and a sequence encoding an activity
CC inducing domain of a second protein. The fusion proteins are useful
CC for the specific killing of tumour cells or the killing of cells
CC infected with intracellular pathogens, especially HIV.
XX
SQ Sequence 776 AA:

Query Match 96.1%; Score 3983; DB 15; Length 776;
Best Local Similarity 100.0%; Pred. No. 2.2e-235;
Matches 776; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGGHGVGMHVHVEKEKNKKNKDEKRNKTOEHLKETIMKHIVKIEVKGEEAVKKEAE 93

```

Db      1 agghgdvgmhvkexkdenkrkdeernktgeehlkemlkhvklvkgaeavkkaeae 60
OY      94 KLEEKVPSDVLWMYKKAIGGKIYIVDGDITRHSLEALSEKKKIKIDYGCDALEHHYV 153
Db      61 kllkxpsdvlwmYkaiygkIyIvdgdltkhislealsedkkkikldiygkdallhehyv 120
OY      154 AKEGYEPVLVIOSESDYVENTEKALANYVEIGKILSRDLSKINOPQKFLDVLTNTKNA 213
Db      121 akegyepvLVIGSSEDYVENTEKALINYYEIGKILSRDLSKINQPYQKFLDVLTNTKNA 180
OY      214 SDSGDGLFTNLQKHPDTFSVEEFLQNSNEQVEWFAKAFAYIIEPQHRDVQLYAPEA 273
Db      181 sdsdgdglftnlqkhpdtfsvefleqnsnevegfakafayIiepqhrdvqlYapea 240
OY      274 FNYMDKFNQOEITLSBELKDQMLSKYEKWEKIKOHYQHSWSLSSEEGGLKKLOIPT 333
Db      241 fnymdkfneqelnlsleelkdgqfmlsryekwekIkqhyqwsdsIsseegfyllkklqipl 300
OY      334 EPKKDITHSLSQEEKELKRIQIDSSDFLSTBEKEFLKKLOIDIRDSLSEEEKELNRI 393
Db      301 epkddilhsIsqeekellkrlqIdssdfIssteekellkldIdIsdsseeekellnrl 360
OY      394 QVDSNPLSEKEREFLKKLKDIOPYDINQRLQDTGGLDPSPINLDVRKQYKRDIONID 453
Db      361 qvdsnplsekekeflkklkldIdpydIngrlqdtgglIdpsInldvrkYkrdIdqnd 420
OY      454 ALLHOSIGSTLYNKIYLYEMKMNINNTATIGALVSDTQNTKIRGIFNFKKNEKYSIS 513
Db      421 allhgisgstlYnkIylyemmnlnlTatlgadlvdstdnkInrgIfnefkknfkYsis 480
OY      514 SNMYAIDINERPALDNERLKWRIQDSPTRAGYLENGKILQINIGELIRDVQIIOKSEK 573
Db      481 snymyidinerpalDnerlkwrlqIsptcragylengkIlIqngIelkdvqllkqsek 540
OY      574 EYRIDAKVVPKSIDTKIQEAOLNINQEWNKALGLPKYTKLITFVNYHNRYASNIYESAY 633
Db      541 eyrIdakvvpkSIkdtkIqeaqnlInqewnkalgIpkykklItfvnhyrnaSniYesay 600
OY      634 LLINEKKNNTOSDLIKKYNVYLVDSNGRPFYDTITPLNTEQOYTRHODEIYQVHSHKLYV 693
Db      601 llinekknlnqsdllkKvkvtyIvdngnrtvfdltlPnlaeqYchdeIyegvhsKglYv 660
OY      694 PESRSILHGPSKGVLELRNDSEGFIEHFGHADVAGYLLDXKQNSDLYVTSKKPFIDIFKE 753
Db      661 pestrIlHgpsKgvlelnDsegfIefghnavdYagylldKngsdlvtnskkfIdIfke 720
OY      754 EGSNLTSGRTNDEEFAEAFRLMHSSTDHAERLKVQKNAPKTFQFINDQIKFTIINS 809
Db      721 egSnltsgYrtneaeffaeafRlmhsstdhaerlKvqknApktfQfIndqIkftIIns 776

RESULT 3
AAU00222
ID      AAU00222 standard; Protein: 485 AA.
AC      AAU00222;
XX
DT      31-MAY-2001 (first entry)
XX
XX      Lfn-Bcl-XL apoptosis-modifying fusion protein.
DE
XX
KW      Human: Lfn-Bcl-XL; apoptosis: cancer; spinal muscular atrophy;
KW      anthrax lethal factor; neoplasm; tumor; hyper-proliferation;
KW      Alzheimer's disease; neurodegenerative disorder; stroke;
KW      transient ischaemic neuronal injury; spinal cord injury;
KW      Huntington's disease.
XX
XX      Chimeric - Homo sapiens.
OS      Chimeric - Corynebacterium diptheriae.
OS      Chimeric - Synthetic.
XX
FH      key      Location/Qualifiers

```

```

FT      Region      5..10
FT      /note="6x histidine tag"
FT      Region      21..276
FT      /note="Anthrax lethal factor amino acids 1 to 255"
FT      Region      277..485
FT      /note="Bcl-XL amino acids 1 to 209"
XX
PN      WO200112661-A2.
XX
PD      22-FEB-2001.
XX
PF      15-AUG-2000; 2000MO-US22293.
XX
PR      16-AUG-1999; 99US-0149220.
XX
PA      (HARD ) HARVARD COLLEGE.
PA      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI      Youle RJ, Liu X, Collier RJ;
XX
DR      WPI; 2001-218343/22.
DR      N-PSDB; AAS00250.
XX
PT      Novel fusion protein for modifying apoptosis in target cell and
PT      reducing apoptosis after transient ischaemic neuronal injury, has two
PT      domains which targets protein to a cell and modifies apoptotic response
PT      of cell
XX
PS      Claim 4; Page 64-65; 65pp; English.
XX
CC      The sequence represents the amino acid sequence of Lfn-Bcl-XL apoptosis-
CC      modifying fusion protein comprising anthrax lethal factor (LF) sequence
CC      fused to Bcl-XL. The functional apoptosis-modifying fusion protein is
CC      capable of binding a target cell and integrating into or crossing a
CC      cellular membrane of the target cell. The apoptosis-modifying fusion
CC      protein comprises at least two domains: the DFR domain, which targets
CC      the fusion protein to the target cell and the Bcl-XL domain, which
CC      modifies an apoptotic response of the target cell. The fusion protein is
CC      useful for modifying (inhibiting or enhancing) apoptosis in a target
CC      cell, such as neuron, lymphocyte, cancer, neoplasm, macrophage.
CC      epithelial, stem, tumor or hyper-proliferative cell or an adipocyte. It
CC      is also useful for reducing apoptosis in a subject after transient
CC      ischaemic neuronal injury, especially spinal cord injury. The fusion
CC      protein may be used to treat various diseases and injury conditions
CC      through inhibition or enhancement of apoptotic cellular response,
CC      including neurodegenerative disorders such as Alzheimer's disease,
CC      Huntington's disease, spinal muscular atrophy, stroke episodes and
CC      unregulated cell growth as in tumors and various cancers. The apoptosis-
CC      modifying fusion protein can be delivered effectively throughout the body
CC      and targeted to selective tissue and cells.
XX
SQ      Sequence      485 AA:

Query Match      31.9%; Score 1322; DB 22; Length 485;
Best Local Similarity 58.8%; Pred. No. 4,7e-73;
Matches 293; Conservative 30; Mismatches 49; Indels 126; Gaps 11;

```


CC components. 508 AA;
XX
SQ

Query Match	31.5%	Score 1307;	DB 15;	Length 508;
Best Local Similarity	99.6%	Pred. No. 4.1e-72;		
Matches 254; Conservative	1;	Mismatches 0;	Indels 0;	Gaps 0

Qy	34	AGGAGDVGMAHKEKKEKNKDEKRRDEERNKTOEHNLEIKMKHIVKIEVKEEYVKKAAE	93
Db	1	aggrgdvgmhyvkekekhkdkerkrdeernktgeehlkelmhkvklevkgeayvkkkaae	60
Qy	94	KLEKRVPSDVLKWKKAIGKRIYVDGDTYKHISLEALSDEKKKIKDITYGKALLNHHYU	153
Db	61	kllkervpsdvlewmkkaigkriyivdgdtlkhlslealsedkkkkikdiygrkdallnhyuy	120
Qy	154	AKEGEPLVYQSSDEVNTEKALANYEELGKILISDILSKINOPQRKLDVLTNTKNA	213
Db	121	akegeyepvlyqssedvyenkekainyyelgkllistdlstklnpqrkldvltntkna	180
Qy	214	SDSDGDLLEFTNOLKEHPTDTSVEFLBONSNEVOEYAKAFAYITEPHKRDVLOQLYAPEA	273
Db	181	sdsdgdqllftfnqlkehpdtstveflbqnsnevgvrtakafayitephqrdvliqlyapea	240
Qy	274	FNYMDKRENEDEINLS	288
Db	241	fnymdkrfnegeinlt	255

RESULT 6

ID AAR60180 standard; Protein; 456 AA.

AC AAR60180;

DT 04-APR-1995 (first entry)

DE LF(1-254) --TR--PE(401-602) toxin fusion protein.

KM Anthrax *Bacillus anthracis*; fusion protein; lethal factor;
 KM Protective antigen; cell killing; targeting; pathogen;
 KM Intracellular; HIV; human immunodeficiency virus;
 KM *Pseudomonas*; exotoxin.
 XX
 OS *Bacillus anthracis*.
 OS *Pseudomonas* sp.

PN WO9418332-A.

PD 18-AUG-1994.

PF 14-FEB-1994; 94WO-US01624.

PR 12-FEB-1993; 93US-0021601.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Arora N, Klimpel K, Leppä SH, Nichols PJ, Singh Y;

DR WPI; 1994-279753/34.

DR N-PSDB; AAR60180.

PT Nucleic acid encoding anthrax toxin fusion protein - useful for targeting toxin to specific cells, eg for killing tumour cells
PT or HIV-infected cells

PS Claim 7; Page 86-87; 124pp; English

CC This sequence is a fusion protein comprising amino acid residues
CC 1-254 of the anthrax protective antigen binding domain of the
CC native anthrax lethal factor, a two residue linker and residues

CC 401-602 of a *Pseudomonas* exotoxin A activity inducing domain of a
CC second protein. Such toxin fusion proteins may be useful for the
CC specific killing of tumour cells or the killing of cells infected
CC with intracellular pathogens, especially HIV, depending on their
CC components.

Sequence 456 AA;

Query Match	31.5%	Score 1306;	DB 15;	Length 456;
Best Local Similarity	100.0%	Pred. No. 4,1e-72;		
Matches 254;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Oy	34	AGGHGDVMAHYKKEKKNDEKRRKDEENKQOEHLKEIMKHIVKIEVKGSEAVKKEAAE	93
Db	1	agghgdvmahykkekknkdenkkrkdeernktgeehlkeimkhivkievkgeevkkaae	60
Oy	94	KLLEKVPSPDLMEYKKAIGKGIYIVDGDITKHSLEALSSEDKKKIKIDYGDALLHENYU	153
Db	61	kllkvpvspdlmeykkaigkgyivdgditkhslealssedkkkikidylgdaallhenyuy	120
Oy	154	AKEGEPEVLVIOSSDDVYENETKRLANYEHLKILTSIDILSKINOPKOFIDVNTKNA	213
Db	121	akegepevlviovssddvyeneteakalnvyeigklltsidilsklnpqrkildvntlkna	180
Oy	214	SDSDGODLLFTNOLKEHPTDSVFEFLBONSNEVOEPAKAFAYVIEQOHRDVLQLVAPEA	273
Db	181	sdsdgdqdllftnqlkehpdtsvfevlbqnsnevgevfakafayvievqrhndvllqlyapea	240
Oy	274	FNMYDKFNEOEINL	287
Db	241	fnymdkfnegeinl	254

RESULT 7

ID AAR04236 standard; protein; 800 AA

AC AAR04236;

DT 12-SEP-1989 (first entry)

DE Adenyl cyclase gene of *Bacillus anthracis*.

adenyl cyclase; pertussis; protective vaccines; signal sequence

PN EP366550-A.

02-MAY-1990

PF 25-OCT-1989; 89EP-0402949.

PR 25-OCT-1988; 88FR-0013952.

PA (INSP) INSTITUT PASTEUR.

PI Escuyer V, Duflot E, Mock M, Danchin A.

DR WPI; 1990-133988/18.

DR N-NSDB; Q04123.

PT Nucleotide sequence encoding adenyl cyclase of *Bacillus anthracis*
PT and derived proteins, useful in protective vaccines, also effective
PT against pertussis

PS Claim 8; 13; 23pp; French.

CC In vivo the adenylyl cyclase protein is synthesised as a precursor with a
CC signal sequence. The mature protein is secreted into the periplasmic
CC space, the signal peptide having been cleaved off at the moment of
CC secretion.


```
Db 377 -----lkvremdlekrehnflh-medqldklxnsfvknn--nqlkvykce--- 418
OY 160 PVLIOSSDYVENTERKALNVVYEIGKILISROILSKINQPYOKFLDYLVNTIKRNASDSQG 219
Db 419 ----lknltelkekkelk---dienvskeeknllnqnekekqlafnkhke--- 468
OY 220 DLLFTNQLKEHPTDFSVFELEQNSNEVOEYFAKAFAYTIEPQHRDYLOLYAPAFVYMDK 279
Db 469 ----lhglkeelke-svktktletqelgem-----vdikkekldgl-----gek 507
OY 280 FNEQELNLSLEBKDOQRMLSRV-----EKWEKIKOHYOHWSLSLEE----- 321
Db 508 ynaqlesistelskkekynkyntylleelnlnnekleeenkytlnqnytnelmlnm 567
OY 322 ----GKGLKKIKQPIEPKKDDI-----IHSIOSEKEL---LKKIQIDSSPFLSTEEL 368
Db 568 dlhmlnqnlkntmgtstlknvdhllneqldklnnekgtlnsklselnlvqimdl--keek 625
OY 369 EFLKKIQIDIRDSLESEKELNRIQVDSSNPLSEKEKEFLKKL----- 414
Db 626 dflnqnlvqlnql-----dltrkmeekenkmllegenkykgemellrgnksesennlm 680
OY 415 DIQPIDINQRL-----QDTGGLIDSPSINLDVRKQYKRDIQNDALLHQS 459
Db 681 deevcdlkrklksesemkmmkeehdkllaekldcdvtr--tremeknekdlnmikee- 737
OY 460 IGSITLXNKIYLVENNMNINLTATIGADLYDSTDNTR-IRGRIFNERKKNNKYSISS--- 514
Db 738 -----yed-kin-----tlkeqnekdlnltlkeqnekdlnltlkeyekhlnmkee 781
OY 515 -NYMIVDINERPALD-----NERLKWRIQISPDTRAGYL-----ENCK 551
Db 782 yehklnltlnegnekhlnltlnegnekhlnmkeeyekdmnltlnegnedkmmksikeeyenk 841
OY 552 LILORNGILEIDV--QIIKQSEKEYIRIDAKVVPKSKIDTKIOEQLNINO----- 601
Db 842 nqlnsnneklkldvnyneyleevdklrvltidek---kkqldkelnyahlikakekeqlllte 898
OY 602 -EMNKALGPKTKLTITFVHNRYASNIYESAVLIIN-EMKNNIQS-DLTKKTYNLVDS 658
Db 899 meelkqrdnkysdl-----yekyl-kllkslcmllnllecddlenedlitrleeyinn 952
OY 659 NGRFVETDITLERNIAQOYTHODEIYEQVHSGLYVPSRSILHGPSKGVLELNDESGFI 718
Db 953 kgl-----kveeekenhkr-----hsstnlkskxf- 979
OY 719 HEPGHAVDYAGYLDKNOSDLVYNSKKF-----IDIFKEGGSNL 758
Db 980 --fksniedekshelkkkkekdlkskdeleeknklikelndklklqdellykkq-sna 1036
OY 759 TSYGRFNEAEFFAEARLMLHSTDHAERLKVOKNAPRTFOFINOIKFI 806
Db 1037 qqvdhkkkswlllkdkskelkdkenglnvekeekdlkkkddetlil 1084
```

```
RESULT 9
AAM39097
ID AAM39097 standard; Protein; 2663 AA.
```

```
XX AAM39097;
```

```
DT 22-OCT-2001 (first entry)
```

```
DE Human polypeptide SEQ ID NO 2242.
```

```
XX Human; nootropic; immunosuppressant; cytosstatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
```

```
XX Homo sapiens.
```

```
XX
PN MO200153312-A1.
PD 26-JUL-2001.
PE 26-DEC-2000; 2000WO-US34263.
PX
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSQ INC.
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RJ;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AA158253.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 4; SEQ ID NO 2242; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AA42213) with nootropic,
XX immunosuppressant and cytosstatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: immune system suppression,
XX activin/inhibin activity, chemotactic/chemokine activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
SQ Sequence 2663 AA;
```

```
Query Match 5.7%; Score 235.5; DB 22; Length 2663;
Best Local Similarity 20.5%; Pred. No. 1e-05;
Matches 178; Conservative 155; Mismatches 305; Indels 231; Gaps 41;
```

```
OY 42 MHVKEKKNK-----DENKRKDEERNKQOE-EHLKEIKKH-----YKREV 81
Db 524 melklkekndldefealerlttkkdgemqlheisnlknlvkhveyngdlenelsskvel 563
OY 82 KGEBAVKKAEAEKULEKVPDLEMYKAIGKITYVDGDTYKHISLEALSEDKKKIK--- 138
Db 584 lre---kedqklklqeyldsqklienlkm-----dis--yslesi-edbpqmkqtl 627
OY 139 -----DIYGDALLHEHYVYAKREYEVLYIOSESDYVENTERKALNVYIEIGKILSR 190
Db 628 fdaetvaldakresafitsenlelkekme---latykgmendilqlygsqleakkmqy 684
OY 191 DISIKTIQPOKFLDYLVNTIKNASDSG---QDLFTNQLKEHPTDFSVF---LEQNSN 244
Db 685 dlekeqsaenleltksll-----dgkypkdlclnlelegklttdlqkelnlveeneea 738
OY 245 EVOGVF---AKAFAYTIEPQHRDV-----LOLYAPEAFVYMDK---NEQELNLS 288
Db 739 lreevlllseklpseverlrlkelqdkseelhltse---kdklfsevnkstrvqgl 794
```

Oy	289	LEEL	---	KDQRMISRYKMKKRIKOHQHMDSLSSEGRLLKLLQIPIEPKKDDIHS-	343
Db	795	leeqtkctdd-	lactgnsyxtqdeqtnf-----	-----kllmndteqykxmwleene	841
Oy	344	-LSOEKELLKRIO	-IDSS-DFLSTE-----	-----EKEFLKKLQIDIRD	380
Db	842	rmngevlnlsaeqkfdfs	lgalktelsyktelqgelktrevgerlnemegkle-q	lenrd	900
Oy	381	S---	LSSEKELLNRIOVDSSNPLSEKKEFEFLKLLDIOPIPDINORLODT-----	428	
Db	901	sp1gtvterekl	-----ltek1qgtleevxtcltqgeddklqjesiqierdqk	949	
Oy	429	GGLIDPSINIDVKKOYKRRDIONIDLHOSIGSTLYNKTYILEMNNININNLATIGADLV	488		
Db	950	sdlnctvmnmldtqeqqltnalesikq-	hgetlnclksfxisevsrnlh-----me	998	
Oy	489	DSTDNKTINRGIFENFEFKNKRYSSISNMYMVIDINERPALDNERLKWRIOLSPDTRAGYLE	548		
Db	999	entgetk-----	defgqk-----mvgldkkqgdleahn-----	10255	
Oy	549	NGKLLIORNIGLETKDVOITIKOSEKEYIRIDAKVVPKSKITDKIOE-AQLNINDEMNKL	607		
Db	1026	-----	qgtltdadvkdnelleqgkrlfsl1qgelneqlgmlesvlaeqkltdlken--	10767	
Oy	608	GLPYTKLIFPNYNNRNASIVESAVLINEMKNINOSLLKVNPNYLVLDGGRVFPDI	667		
Db	1077	-----	lemctlenq-----eelrl1lgelkk--qgelvqeqekhnakkge1stcd	11200	
Oy	668	TLPNIAEQYTHODEIYEOVHSKGLVYPESKSTLLHGPSKVELRNDSEGFINEFGHAAYD	727		
Db	1121	rlaeveek1kexsqgqkqeqql1nvqemsemqkk1nelen1knelkkel1tlemete	1180		
Oy	728	--YAGYLLDKNOSL--VTNSKRFI----	DIREESNLTSGRINAEAFPEARL--	776	
Db	1181	rltlaqk1nemyeevks1tker1vklkelqk1steterdhlrgyl1re1eatg1qtk1eelkia	1240		
Oy	777	-MSTDAERLKVOKN--APKTFQFINDO	802		
Db	1241	h1h1kehgetdelr1rsvsektq1lntq	1269		
RESULT 10					
AA	40883	standard; Protein; 2688 AA.			
AC	40883;				
DT	22-OCT-2001	(first entry)			
DE	Human polypeptide SEQ ID NO 5814.				
KW	Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;				
KW	peripheral nervous system; neuropathy; central nervous system; CNS;				
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;				
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;				
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;				
KW	leukaemia.				
OS	Homo sapiens.				
PN	WO200153312-A1.				
PD	26-JUL-2001.				
PF	26-DEC-2000; 2000MO-US34263.				
PR	21-JAN-2000; 2000US-0488725.				
PR	25-APR-2000; 2000US-0552317.				
PR	09-JUL-2000; 2000US-0598042.				
PR	19-JUL-2000; 2000US-0620312.				
PR	03-AUG-2000; 2000US-0653450.				
PR	14-SEP-2000; 2000US-0662191.				

```

PR      19-OCT-2000; 2000US-0693036.
PR      29-NOV-2000; 2000US-0727344.
XX
XX
PA      (HYSE-) HYSEQ INC.
XX
XX      Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI      Wang Z, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI      Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX      WPI: 2001-442253/47.
DR      N-PsDB; AA160039.
XX
XX      Novel nucleic acids and polypeptides, useful for treating disorders
PT      such as central nervous system injuries -
XX
PS      Example 2: SEQ ID NO 5814; 10078bp; English.
XX
XX      The invention relates to human nucleic acids (AA157798-AA161369) and
CC      the encoded polypeptides (AAM36642-AAM442213) with nootropic,
CC      immunosuppressant and cytostatic activity. The polynucleotides are useful
CC      in gene therapy. A composition containing a polypeptide or polynucleotide
CC      of the invention may be used to treat diseases of the peripheral nervous
CC      system, such as peripheral nervous injuries, peripheral neuropathy and
CC      localised neuropathies and central nervous system diseases, such as
CC      Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC      lateral sclerosis and Shy-Drager Syndrome. Other uses include the
CC      utilisation of the activities such as: Immune system suppression,
CC      Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC      and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC      assays for receptor activity, arthritis and inflammation, leukaemias and
CC      C.N.S disorders.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification.
XX
XX      Sequence 2668 AA:
SQ

```

OY	429	GGLIDSPSTINDVRRQYKRDIONDALLHLSIGSTLYNKKIYLYEKNMINNTATIGADLV	488
Db	974	sfshdtvmmidltqeglinaelslq--hgetintlksiseevsnln-----me	1022
OY	489	DSTDNTKINRGIFNEFKKFKYSSISNMYIVDINERPALDNERLKWRIQLSPDTRAGYLE	548
Db	1023	entgact-----defgk-----mvgldkkqgleakn-----	1049
OY	549	NCKLLQNRNIGLEIDVOQIKOSEKEYIRIDAKVVPKSKRIDPKIOE-AQNLNIOENKAL	607
Db	1030	-----tqltadvkdnelleqgrkflfsllqeknelqmlsvlaekqldkldken---	1100
OY	608	GLPKYTKLTFFNVNHRYSNIVESAYLILNEKNKNIOSDLIKKVTNYLVDGNGRFVFTDI	667
Db	1101	-----lemtenq-----eairllgdclkt--qgelvageknhalksgegstcd	1144
OY	668	TLPNIAEQYTHODEIYEOVHSGKLYVPESRSLILHGPKSGVELRNDSEGFTHFGHAVDD	727
Db	1145	rlaaveeklkexsqldqeqgdlhvvgeemsemqkkinetienlkhelkkellemete	1204
OY	728	--YAGYLLDKNOSDL--VYNSKKFI---DIRKEGSSNTSYSGRTNEAFPAAPRL---	776
Db	1205	rlqlqgklnenyevksitkerkvklkelqsfeterdhrlrylreiaeatgltkeelkia	1264
OY	777	-MSTDDHAERLKVKQRN--APKTFQETINDQ	802
Db	1265	hhlkhegetidelrrsvsektagqlntq	1293
RESULT	11		
AAW24790			
ID	AAW24790	standard; Protein; 1786 AA.	
XX			
AC	AAW24790;		
XX			
DT	08-OCT-1997	(first entry)	
XX			
DE	P. falciparum	liver stage antigen-3.	
XX			
KW	Plasmodium falciparum; pre-erythrocyte; liver stage antigen; serum;		
KW	prophylaxis; That strain; gene organisation; exon; intron; hydrophobic;		
KV	glycosyl-1-phosphatidylinositol membrane anchoring sequence; antibody;		
KW	vaccine; immunotherapy; malaria.		
XX			
OS	Plasmodium falciparum.		
XX			
FH	Key	Location/Qualifiers	
FT	Region	223..278	
FT		/note="repeat region 1"	
FT	Region	279..818	
FT		/note="repeat region 2"	
FT	Region	1537..1576	
FT		/note="repeat region 3"	
XX			
PN	MO9641877-A2.		
XX			
PD	27-DEC-1996.		
XX			
PF	12-JUN-1996;	96WO-FR00894.	
XX			
PR	13-JUN-1995;	95FR-0007007.	
XX			
PA	(INSP) INST PASTEUR.		
XX			
PI	Daubersies P, Druilhe P;		
XX			
DR	WPI: 1997-065464/06.		
DR	N-PSDB; AAT78868.		
XX			
PT	Plasmodium falciparum poly.peptide(s) and related nucleic acids -		
PT	derived from the liver stage antigen-3, useful for malaria vaccine		
PT	prodn. and diagnosis		
XX			

PS Claim 1, Fig 2A-I; 69pp; French.

XX This sequence corresponds to a Plasmodium falciparum strain K1
CC pre-erythrocytic liver stage antigen-3 (LSA-3) protein. The encoding
CC gene sequence was isolated by screening a P. falciparum strain 79/96
CC cDNA library with serum from a missionary treated by prophylaxis (for strain
CC T6/96 see F90101286). Of 20 clones isolated, clone 7295 was used to
CC screen a library generated from Thai strain K1. One clone contained a
CC 6.85 kb insert including the genomic sequence AAT78867. The gene
CC comprises a 1.8 kb region encoding 3 major blocks of tetrapeptide
CC repeats (especially the amino acid sequence VEES, VEEN, VEEL, VAPS, VAPT
CC etc) and a 3 hydrophobic region corresponding to a
CC glycosyl-phosphatidyl- inositol membrane anchoring sequence. The
CC invention relates to new polypeptides of at least 10 amino acids derived
CC from the LSA-3 protein with the exception of the peptides AAW24791-4.
CC The LSA-3 peptides can be used to raise antibodies and as vaccines for
CC immunotherapy of malaria.

XX
XQ Sequence 1786 AA:

Query Match	5.5%;	Score 228.5;	DB 18;	Length 1786;
Best Local Similarity	18.9%;	Pred. No. 1.6e-05;		
Matches 155;	Conservative 134;	Mismatches 271;	Indels 259;	Gaps 32

Oy	44	VKEEKKKKDKRRDEBRRNTOEHHKELMKHIVKLEVGEAAVKEAAKLEKVP--	100
Db	891	leekleelnehnysaalenqseekevdyt-----eevkeevattileetvegae	942
Oy	101	-----SDVLEMKKAGKGVY--YVGG-DITVHISLEALSE--	132
Db	943	eksantlletfelnleenaavemenaveneleklinevforvldkveetveisgeslenmem	1002
Oy	133	DKKIKIOIKYKDALHHEHYVYAR--EGYEPVLYIGSSE-----DYVENTEKAL	178
Db	1003	dkafissetfnvxygigenlltgmfrsietisyigseekevdlneuvssalldnenmkegl	1062
Oy	179	-----NVYEYIGKILSRD-----ILSKINOPYOKFLDLVNLNTK	211
Db	1063	lnklienstsegyetvethvegnvydvdpymkdgflgllneegylknefnfinedfk	1122
Oy	212	NASSDODDLFTFNOLKEMHTDTSVEFLEONSNEQVYFAKFAFYIEPQHRDVLQLYAR	271
Db	1123	ses-----dvlveealkdep-----vqkeketvallleemeenlvdvlee	1163
Oy	272	EAFNVYMKF-----NEOEINTLEELTKDQMLSRYEKWEKIK	308
Db	1164	ekelltkmldaveeslejsdskeeteslkdkexkdvslyveevdndmdesvekyvlejk	1223
Oy	309	QHYQHW-----SDSLSEGRGL-----LKRIQPIE-----PKKD	338
Db	1224	nmeemlkkdavehndltfslkleetqelneveadllkdmeklkhelekalsedskeiidakd	1283
Oy	339	DIHLSLQOEKEL-----LKRIQIDS-----SDPLSTEE-----KEF-----	370
Db	1284	dlekvleeneidlttcldevelldkveeekiekvsldkledllkeykeikelleselle	1343
Oy	371	-----LKRIQIDIRDSLS-----EEBEKELNRIOVDSNSPLSEKKEFELKKLKLDI	416
Db	1344	dykelktleedllleekkelekhfkeeeaaekldleedllkevsileveeekle---	1400
Oy	417	QPYINORLOD-----TGLIDSPSINLDVKYOKRKRIQINDA	454
Db	1401	evhalkveehllsagdahlkgleeddlleevddlksgilmdlkymdlgmdekstedtvc	1460
Oy	455	LLHOSIGSTLYNRYLYEENNINNLATGADLVOSTDNTKIRNG-----IFNFEKKNF	508
Db	1461	kligerveysk-----dvlsaalgmdeeqmkttkkqqrpxlleevlllleevveep	1508
Oy	509	KYSISSNMYIVDINERPAID-----NBLKWRIO--LSFDRAGVLENGKLLIORNT	558
Db	1509	kkkltktkkvrlkdkpkelvevmekdedleedvveedleedleadvdeiddeid	1568

Query Match	5.4%	Score 224.5	DB 21	Length 1558
Oy	558	GLEIFD--VOIKOSKEKIRIDAKVPPSKIDTKIOEAOINQENKALGLPKRYIKLI	616	
Db	1569	l-ldkdevldllivqkxriekvak---kklekkeevgsjlkhdvym---kyqki	1621	
Oy	617	TENVHNRVASNIVESAYLLIENKKNIOQLIKKYNLY	655	
Db	1622	dkev-dkevskakeskndvtnlkgp--gdffskvknfv	1657	
RESULT 12				
ID	AA18324	AA18324		
XX	AA18324	standard; Protein; 1558 AA.		
AC	AA18324			
XX	AA18324			
DT	07-NOV-2000	(first entry)		
XX				
DE	Plasmodium falciparum chromosome 2 related protein SEQ ID NO:182.			
KW	Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;			
RV	antimalarial; malaria; protozoacide; infection; insecticide.			
XX				
OS	Plasmodium falciparum.			
PN	WO200025728-A2.			
PD	11-MAY-2000.			
PF	05-NOV-1999; 99WO-US26796.			
PR	05-NOV-1998; 98US-0107131.			
XX				
PA	(HOEF/) HOFFMAN S.			
PA	(CARU/) CARUCCI D.			
PA	(GARD/) GARDNER M.			
PA	(VENT/) VENTER J C.			
XX				
PI	Hoffman S, Carnucci D, Gardner M, Venter JC;			
DR	WPI: 2000-365347/31.			
XX				
PT	Proteins encoded by chromosome 2 of the human malarial parasite,			
PT	Plasmodium falciparum, useful as antimalarial vaccines and in the			
PT	diagnosis of P.falciparum infection -			
XX				
PS	Disclosure: Page 410-414; 577pp; English.			
XX				
CC	The present invention describes proteins and their fragments (I) encoded			
CC	by chromosome 2 of the human malarial parasite, Plasmodium falciparum.			
CC	Also described are: (1) nucleotide sequences (II) encoding (I); and (2)			
CC	vaccines against P. falciparum infection comprising (I) or (II).			
CC	(I) and (II) are useful for the development of vaccines against			
CC	P. falciparum infection. (I) and polyclonal antisera or a monoclonal			
CC	antibody raised to immunogens comprising the sequences of (I) are			
CC	useful in the detection of infection with P. falciparum. Furthermore,			
CC	(I) (especially when they are rifins or secreted or membrane proteins)			
CC	can aid the identification of drugs to treat or prevent P. falciparum			
CC	infection, or they can be used to identify drug resistance in			
CC	P. falciparum. Sequencing of the Plasmodium chromosome 2 and the			
CC	subsequent identification of proteins encoded by it will help to expand			
CC	our understanding of parasite biology, a process hampered by the			
CC	complexity of the parasite life cycle, and provide new targets for			
CC	vaccine and drug development. Parasite resistance to drugs and mosquito			
CC	resistance to insecticides have led to a resurgence of malaria in new			
CC	parts of the world, and there is a pressing need for vaccines and many			
CC	drugs. AA70078 to AA70287 and AB18144 to AB18352 represent nucleotide			
CC	and protein sequences given in the present invention, but which are not			
CC	specifically mentioned within the specification.			
XX				
XX	Sequence 1558 AA;			

Best Local Similarity 18.5%: Pred. No. 2,4e-05:
Matches 156; Conservative 123; Mismatches 260; Indels 305; Gaps 29.

```
QY      44 VKEKKNKKDEENRKRTOEHLKELMKHVKLEVGGEVKKAEAKLEKV----- 99
Db      659 IEEKLLENVLsaalenltgseeekveidyI-----eevkeavtllietvegqe 710
QY      100 ---PSDLEMRYAIGGKIIVDGDIRKH-----STELASE 132
Db      711 eesastlfeleenavaesnenvenaelklnetvfntvlkveetvelsgselemnm 770
QY      133 DKKKIDIRGKDALLHENHYVAK-EGYEPLVIQSSE-----DYVENTEKAL 178
Db      771 dkafaeildnkvqgenlltgmfirsletsivlygseekvdlnenvsslldnlennkegl 830
QY      179 -----NYYVELGIKLISRP-----tlSKINQPQKFLDYLVNTIK 211
Db      831 lklhenisstegvetvehvegnvydvdpamkqflgllneagglkemfnfledvfk 890
QY      212 NASDSGDODLLFTNLQLEHPDTFSVEFLQNSNEVOEFKAFAAYIEPQHRRVDQLAVAP 271
Db      891 ses-----dvtiveelkdcp-----vqkevketvsllleemeeniydvlee 931
QY      272 EAFNYMDK-----NEGTINLSLELKDQRMLSRYEKKETK 308
Db      932 ekreditdkmidaveesielsdskeeetaskdekqvsliveevgdmdndesvekyletk 991
QY      309 OHXOHW-----SDSLSEGRGL-----LKKLOPIPE-----PKKD 338
Db      992 nmeeelmkavaelnditsklieetgeqlneaveadlkdmekllelekalsedskelidakk 1051
QY      339 DIIHLSOEKEL-----LKRIOIDS-----SDFLSTE-----KEF----- 370
Db      1052 dclekveiehdtlttledevelkdveedkleksdlkdleedllkvekkeleeseile 1111
QY      371 ----LKKLOIDRDSLSEKELRNIRIQVSSNPULSEKEKFKLKKLKDIOFYDINQRQ 426
Db      1112 dykelktletdilleekkeiekdhfef-----eeeaeklleaadi-----LK 1154
QY      427 DNGGLDSPSINDVAKOKRKDIOMT-----DAL-----LHGSGSTLYNKIY 469
Db      1155 evssleveeekkleevehelleevenhlsgdahnkgleeddleeovdlkgslldmkygone 1214
QY      470 L--YENNMINNLTAFLGADLVLDSTONTKINRGIFNEFFKFNKYSSSNMYIWDINERPAL 527
Db      1215 lgdmkhesledvtakgervesalokvlssaalgmedeqmtlrka-----grpl 1263
QY      528 DNERLKWRIQLSPDRAGYLENGKLILORNIGLEIKD-----564
Db      1264 eeallkeevkeep-----kkittkkvfrldlkdkpkelaijevemkedidedlee 1314
QY      565 -----VOIKOSEKKEVRIPDAKKVPKSIDRK 591
Db      1315 dreedleedkvedidedidedidegedkdevidlivkkekrtrekvkck---kkklek 1371
QY      592 IOEAOLINQENMKKALGLPRYTRLTFNVHNRYASNIIVESAYLLINEMKNINIOSDLIKV 651
Db      1372 veegyvglikkhvdevm---kyvgkldkey-dkevskaleskndvtnvlqn--gdfefskv 1425
QY      652 TNVL 655
Db      1426 knfv 1429
```

RESULT 13
AAIB18294
ID AAIB18294 standard; Protein: 980 AA.
XX
AC AAIB18294;
XX
DT 07-NOV-2000 (first entry)
XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:157.

```

XX  Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KM  antimalarial; malaria; protozoacide; infection; insecticide.
XX
XX  Plasmodium falciparum.
XX
XX  WO200025728-A2.
XX
XX  11-MAY-2000.
XX
XX  05-NOV-1999; 99WO-US26796.
XX
XX  05-NOV-1998; 98US-0107131.
XX
XX  (HOFF/) HOFFMAN S.
PA  (CARU/) CARUCCI D.
PA  (GARD/) GARDNER M.
PA  (VENT/) VENTER J C.
XX
XX  Hoffman S, Carucci D, Gardner M, Venter JC:
PI  WPI; 2000-365347/31.
XX
XX  Proteins encoded by chromosome 2 of the human malaria parasite,
PT  Plasmodium falciparum, useful as antimalarial vaccines and in the
PT  diagnosis of P. falciparum infection -
XX
XX  Disclosure; Page 362-365; 577pp; English.
XX
XX  The present invention describes proteins and their fragments (I) encoded
CC  by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC  Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC  vaccines against P. falciparum infection comprising (I) or (II).
CC  (I) and (II) are useful for the development of vaccines against
CC  P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC  antibody raised to immunogens comprising the sequences of (I), are
CC  useful in the detection of infection with P. falciparum. Furthermore,
CC  (I) (especially when they are rifins or secreted or membrane proteins)
CC  can aid the identification of drugs to treat or prevent P. falciparum
CC  infection, or they can be used to identify drug resistance in
CC  P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC  subsequent identification of proteins encoded by it will help to expand
CC  our understanding of parasite biology, a process hampered by the
CC  complexity of the parasitic life cycle, and provide new targets for
CC  vaccine and drug development. Parasite resistance to drugs and mosquito
CC  resistance to insecticides have led to a resurgence of malaria in many
CC  parts of the world, and there is a pressing need for vaccines and new
CC  drugs. AAAY0078 to AAAY0287 and AAB18144 to AAB18352 represent nucleotide
CC  and protein sequences given in the present invention, but which are not
CC  specifically mentioned within the specification.
XX
XX  Sequence 980 AA:
SQ

```

```

Query Match 5.3%; Score 218; DB 21; Length 980;
Best Local Similarity 21.2%; Pred. No. 3.3e-05;
Matches 150; Conservative 126; Mismatches 262; Indels 168; Gaps 32;

```

```

QY 290 --EELKQORML--SRYEKMEKIKQHYQWMSDSLSEBGRGLKKTQIPIEPKDDIHSLS 345
DB 509 kneelnekeiilknydk--elmlleqnykk lqeeqdm lnnk lksndqtlknq- lsemg 565
QY 346 QEEKELLRKIQ-----IDSSDFLSTEKEFLKQIDIRDSLSEEEKELNRIQ-VDS 398
DB 566 eenkelkrlknvcdm lqsg lllkeneklmqkve-eyknl lqkqgelkn lnyqder 624
QY 399 NPLSEKEKEFL-----KKLKDIDQPYDINORLQDTGGLIDSPSINLOVRK 443
DB 625 lclqkemedlvndceeklkqak lnnk lltatmannnn lmden lkekdkkn-clmk 683
QY 444 QYKRDIONIDALL-----HQSIGSTLYNKKIYLYEMKNINNLATGADLVSDTWTKI 496
DB 684 dmeekkeelnlk lveeksk lshv--k lqemsl lveqn-ek lkeemg lsr lal kdmee l 740
QY 497 NRGI--FNEPKKNFKYSISSNVMIVDINERPALDNERLKNRIQLSPPRAGYLENGKIL 554
DB 741 kkmekyeekkkneerkkneerkkneerkkneerkkneekkkne--eerkkneekkklek 798
QY 555 QRNIGLEIKD-VQIIKQSEKEYIRIDAKVVPKS-----KIDTKIQEAQLINQEW- 603
DB 799 dkngfeekkerme lyeqkqdrkrkkkkkghsdkeekyknkkek ltxksn l l fdeey l 858
QY 604 -----NKALGLPKYTKLTFNVHNRYAS-----NIVSATLILNEMW 639
DB 859 lqlleldtgencf lylks lskeldv- llnk lkskdall lndafnkl laltsw l fne- 916
QY 640 KNNIOSDLIRKVTNVLVDNGRFFVFTDTL-PNIAEQYTHODEIYE 684
DB 917 -enkegdn lltventalegn-----l lden ltervemneeyk 954

RESULT 14
AACG82169
ID AACG82169 standard; Protein: 1145 AA.
XX
AC AAG82169;
XX
XX 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:1432.
XX
KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX
XX vaccination; endocarditis.
XX
OS Staphylococcus epidermidis.
XX
XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30782.
XX
XX 09-NOV-1999; 99US-0164258.
XX
PA (GLAX ) GLAXO GROUP LTD.
XX
XX KImmeryly WJ;
XX
XX WPI: 2001-316495/33.
XX
XX N-PSDB: AAH53019.
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
PS Claim 18; Page 405; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
XX (I) and (II) can have antibacterial activity and therefore can be used
XX in vaccination. The nucleic acids (I) may be used to produce the

```

CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH5971 to
CC AAH5970 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH5971 to
CC AAH5978 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4455 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
CC
CC Sequence 1145 AA:

Query Match 5.2%; Score 215; DB 22; Length 1145;
Best Local Similarity 19.2%; Pred. No. 6.1e-05;
Matches 161; Conservative 154; Mismatches 271; Indels 254; Gaps 38;

OY 41 GHHVKEKKNKDKRDEKRNKTOE-----EHLKEIMKRIIVKTEVGEAAVKEAAE 93
DB 361 gfhfkrkpkkeeqikrlneatqgkqyngvgrqprfemsfltrfkn--havnek1ln 418
OY 94 KLEKVPSPDVLBY-----KAIGKRIYVDSGIT-----KHISL 127
DB 419 qeydvvpvsliselqyqtqstlnlyltsdevikalnkl-----enesipfecaavnhqv 475
OY 128 EALSDDKKIKIRIY-----KDAL-----LHEHYVAKESYEPVLIQSSSEDEVENTE 175
DB 476 nelsedenedreydyrlnelnclksltsnhykhyihldslk--ligrtehtfelqg 533
OY 176 KALANYVEIGKILSRDILSKINOPKFLDVL-----NTIKNASDS-----DGO----- 219
DB 534 enstyhnrkhetqnefrnftsgndikraldivkdvplfrtqgdldcllrlndqtkwg 593
OY 220 -----DLFTNQIKHEPTDFSVFLEQNSNEVOEFAVAYIEPQRNDV 265
DB 594 vfgttsagksllnallgenlyvsspnpitaattelsygskeqfclksqeglle-evnhv 652
OY 266 LQLYAPFAFNVDKRNDEGIN-LSLEELKDO-RMLSRVKKMKIKOHQMSDSISEGR 323
DB 653 lqfy-eisfnltdfiesdldkikikleknglaifisalek-----hyemysml----- 700
OY 324 GLLKRLQIPIEPKDDIHSLSQEEKELLRIOIDSPLESEKFLKLIQID----- 377
DB 701 -----ehslhtvsleek-----kwsaedeyatfvktvnhklpdlwlykqi 742
OY 378 IRDSI-----SEEEKELLRIOVDSSNPL-----SEKEREFLKLIQIDQPYDINO 423
DB 743 lldsfghnmgqrhneteqqltsdillvyfyfnhsfndkatiemk-----dmng 796
OY 424 RLQDTGGLDSSIMLDPRKQYKRIQNDALHQSIGS-TLYNKIYLY-----ENM 474
DB 797 lneqgafmllnavdl-----aedkgdigaavedyadalqgvnhselysvsrsgslngml 853
OY 475 NINNTTAT----- 482
DB 854 glnelreslqyfkavesrtlllegqmtylqgmntsfrmlkfhdnaklsarqkhlmy 913
OY 483 -----LGADLVDSDT-----NTKINGIFNEFKKRNKYSISSNMYMVDINE 523
DB 914 knqtrlngeidldtaqrfrneveeqvynherklqlldevsvfnsqmtgn-----dfne 970
OY 524 RPAIDNRLKMRIOQSPOTRACYLENGKLIJOR-----NIGLEIKDVOIIOKESEYIRI 578
DB 971 -----ekristklyldqthqrflfleg-sllterikkylfnsqleeqilpymkknqjnhvll 1024
OY 579 DAKV-VPRSKIDTKIOEQALNINQEMNKALGLPKY-----TKLTTFVNHNRKYSINYESAVL 634

DB 1025 nakfrnepsivdpllqgleln-----smqlsfpkqltkrkynpksqkqigehianqtle 1079
OY 635 ILNENKNNIQ---SDLIKRVNLYLVGNGREFFTDITLPINAE--QYFHODEIYQVMSK 689
DB 1080 llqgdlnsrtrrlnqynhemty-lnehqgmletsig-qgideallsftcltdcllqglek 1137

RESULT 15

AAH18195
ID AAH18195 standard; Protein; 1516 AA.

XX AAH18195;

DT 07-NOV-2000 (first entry)

DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:52.

KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KM antimalarial; malaria; protozoacide; infection; insecticide.

OS Plasmodium falciparum.

PN WO200025728-A2.

XX 11-MAY-2000.

PD 05-NOV-1999; 99WO-US26796.

PF 05-NOV-1998; 98US-0107131.

PA (HOEF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.

PI Hoffman S, Carucci D, Gardner M, Venter JC;

XX WPI: 2000-365347/31.

PT Proteins encoded by chromosome 2 of the human malarial parasite,
PR Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -

PS Disclosure: Page 120-124; 577pp; English.

CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasite lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAH70078 to AAH70287 and AAH18144 to AAH18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.

XX Sequence 1516 AA:

Query Match 5.1%; Score 213; DB 21; Length 1516;
Best Local Similarity 19.8%; Pred. No. 0.00012;

Matches 162; Conservative 136; Mismatches 271; Indels 248; Gaps 37;

```

0Y 45 KEKKK-----KDEKRRKDEENKTOEHELEIM-----KIIIVLEWKGAEVAKKE 90
Db 677 kalsknctnlnlkdylnvlgkkskkkkkflnldlnlynfteeskvgdylgvkkes---ke 734
0Y 91 AAEKLEKV-----PSDVLWYKAIIGCKRYIVDSDITRKHLSLALSBDK 134
Db 735 dtkngdftvgecyrmndlrthcdksdflfnklidnkkkyel---ynltelegelelnkk 791
0Y 135 K-----KIKDYGKDALHHEHYVAKEGYEPVULVIOSSDYVEMTEKALNV 180
Db 792 nynknmdsnktfflknefkkdillldsqifgdslladlkeynytadndnnnksl 851
0Y 181 YVEIGKILSRDILSKINOPQKFLVDVLYMNTKANSAS---DGGDLFTNOLKEHP---TFDS 235
Db 852 yedgenflltr-----nepltneyeeknnllylstedeqkneedlffkkllekeknmts 905
0Y 236 VEPLEONSNEVOEYFAKAFAYYIEPQHRDYLQVAPFAEYNYMDKENEQENLS---LEE 291
Db 906 sddfcnes--vqe-----klynekleeynknkddksssssslllee 945
0Y 292 LKQOR-----MLSRYEKKERIKQNYOHMS-----DSLSEEGRGLLK 328
Db 946 lkyrkkekdelvaspnclvlldefehsndlenmyisvsddmktvnsknlgyvkenvydk 1005
0Y 329 LQIRIEPKDDIILHSOEEKELKRLRIQDSSDFLSTEEKFEFLKRIQIDIRDSSEKE 388
Db 1006 tnveydkkgddgytelsfeshklesktddmnlynddelekhls---kdyisvdkn 1062
0Y 389 LNRIRIYDSSNPLSEKKEFEFLKRLKIDIPYDINORLODTGGLDPSI---NLDVRKQY 445
Db 1063 hvnnl-ynlergederenef-----venklyteshkeneficlenkslkyg 1109
0Y 446 ---KRDIONIDALLHOSIGSLYKLYLENNMINNLTAFLAGD-LVDSDTWTKIKNRGFN 502
Db 1110 mskednsnvll-----ksddlmnskanyfeyllldk-----k 1142
0Y 503 EFKKNFYKSISSNMYIVDINERPALDNERLKRRIQLOSPTRAGYLENCKLILORNGLEI 562
Db 1143 qvmdnfqmglnegu-----ndkllk-edkldgayfeylednkil----- 1179
0Y 563 KDVQIIRIHSKEKEYIRIDAKVVPKSKIDTRYIOEQOLINIDENMKAL-----GLPKTYKL 615
Db 1180 ---dsyilekrlke-----neelikeyk-----klknmlneidemmddkl1lnffgylqisq 1230
0Y 616 ITFVNHRRYASN-----IVESAYIILDEMKNNIOSDLIRKVTYUWLVDGNGRFVFDIT 668
Db 1231 ceeaeqcsylnnkhcdaisdsdsvlvtsgktvlnkfnfkkt----- 1274
0Y 669 LPNIAEOTYHODELIEYO--VHSK-GLYVESRST-LMHGFS---KGVELRNDSEGTIH 719
Db 1275 -----veyekkaaleeklgylygeaelnislscgdcytlvgbhvgjgvale--llk 1322
0Y 720 EFGHAVDVGAGYLLDKMQSLDNLVNSKKFLD--IFKEE 754
Db 1323 aefpnfed-----klklklyvsnpfirkldkmmnye 1352

```

Search completed: December 2, 2001, 13:48:11
Job time: 131 sec

THIS PAGE BLANK (USPTO)


```
Db 1 AGGHQDVGMHVKERKKNDEKRNKQOEHLKIMKHIVKIEVKGGEAVKKEAAE 60
QY 94 KLEKVPDVLDMYKAIGKIYIVDGDITTKHISLEALSSEDKKKIKDIYKGDALLHEHYV 153
Db 61 KLEKVPDVLDMYKAIGKIYIVDGDITTKHISLEALSSEDKKKIKDIYKGDALLHEHYV 120
QY 154 AKEGEPELVYIOSSSDYVENTEKALNVYEYEGKILSRDLSKINOPYOKFLDVLNTIKNA 213
Db 121 AKEGEPELVYIOSSSDYVENTEKALNVYEYEGKILSRDLSKINOPYOKFLDVLNTIKNA 180
QY 214 SDSGQDILLFTNOLKEHPTDSVEFLQNSNMEVOVFKAFAVYIEPOHRYDLQIYAPEA 273
Db 181 SDSGQDILLFTNOLKEHPTDSVEFLQNSNMEVOVFKAFAVYIEPOHRYDLQIYAPEA 240
QY 274 FNYMDKFNQEIINLSLEELKQRMLSRYEKWEKIKQHYQHSDSISEGRGLKKLQIPI 333
Db 241 FNYMDKFNQEIINLSLEELKQRMLSRYEKWEKIKQHYQHSDSISEGRGLKKLQIPI 300
QY 334 EPKDDIITHLSQOEKELKRIQIDSSDFLSTEEKEFLKKIQIDIRDSLSEEEKELMRI 393
Db 301 EPKDDIITHLSQOEKELKRIQIDSSDFLSTEEKEFLKKIQIDIRDSLSEEEKELMRI 360
QY 394 QVDSNPLSEKEKEFLKKLQIDOPYDINORLODPTGGLIDSPSINLDVRYKQYKRDIONID 453
Db 361 QVDSNPLSEKEKEFLKKLQIDOPYDINORLODPTGGLIDSPSINLDVRYKQYKRDIONID 420
QY 454 ALLHOSIGSTLYNKIYLYENNINNL7ATLQADLVSDTNTKINRGITNEKKNEKYSIS 513
Db 421 ALLHOSIGSTLYNKIYLYENNINNL7ATLQADLVSDTNTKINRGITNEKKNEKYSIS 480
QY 514 SNMYIVDINERPALDNEELKRIQISPTTRAGYLENGKLLIQRNIGLEIKVOIITKQSEK 573
Db 481 SNMYIVDINERPALDNEELKRIQISPTTRAGYLENGKLLIQRNIGLEIKVOIITKQSEK 540
QY 574 EYIRIDAVVVKSKIDTQIOEAQNLINQEWKALGLPYTKLITFNVNRRASNVESAY 633
Db 541 EYIRIDAVVVKSKIDTQIOEAQNLINQEWKALGLPYTKLITFNVNRRASNVESAY 600
QY 634 LILNEMKNNIOSDLIKKTYNTLVDSNGRFVFTDITLPIAEOYTHODEIYEQVHSGLYV 693
Db 601 LILNEMKNNIOSDLIKKTYNTLVDSNGRFVFTDITLPIAEOYTHODEIYEQVHSGLYV 660
QY 694 PESRSILHGSQKVELNDSSEGFIERGHAVDVPAGVLLDKNOSDLVTSNKKFLDIPKE 753
Db 661 PESRSILHGSQKVELNDSSEGFIERGHAVDVPAGVLLDKNOSDLVTSNKKFLDIPKE 720
QY 754 EGSNLTSGRTNEAEFFAEFRLMH 778
Db 721 EGSNLTSGRTNEAEFFAEFRLMH 745

RESULT 2
US-08-849B-2
; Sequence 2, Application US/08082849B
; Patent No. 5677274
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Arora, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: Anthrax Fusion Proteins and
; TITLE OF INVENTION: Related Methods
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESS: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/082,849B
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/021,601
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-161-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 776 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-849B-2

Query Match 95.9%; Score 3825; DB 1; Length 776;
Best Local Similarity 100.0%; Pred. No. 3,5e-239;
Matches 745; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGGHQDVGMHVKERKKNDEKRNKQOEHLKIMKHIVKIEVKGGEAVKKEAAE 93
Db 1 AGGHQDVGMHVKERKKNDEKRNKQOEHLKIMKHIVKIEVKGGEAVKKEAAE 60
QY 94 KLEKVPDVLDMYKAIGKIYIVDGDITTKHISLEALSSEDKKKIKDIYKGDALLHEHYV 153
Db 61 KLEKVPDVLDMYKAIGKIYIVDGDITTKHISLEALSSEDKKKIKDIYKGDALLHEHYV 120
QY 154 AKEGEPELVYIOSSSDYVENTEKALNVYEYEGKILSRDLSKINOPYOKFLDVLNTIKNA 213
Db 121 AKEGEPELVYIOSSSDYVENTEKALNVYEYEGKILSRDLSKINOPYOKFLDVLNTIKNA 180
QY 214 SDSGQDILLFTNOLKEHPTDSVEFLQNSNMEVOVFKAFAVYIEPOHRYDLQIYAPEA 273
Db 181 SDSGQDILLFTNOLKEHPTDSVEFLQNSNMEVOVFKAFAVYIEPOHRYDLQIYAPEA 240
QY 274 FNYMDKFNQEIINLSLEELKQRMLSRYEKWEKIKQHYQHSDSISEGRGLKKLQIPI 333
Db 241 FNYMDKFNQEIINLSLEELKQRMLSRYEKWEKIKQHYQHSDSISEGRGLKKLQIPI 300
QY 334 EPKDDIITHLSQOEKELKRIQIDSSDFLSTEEKEFLKKIQIDIRDSLSEEEKELMRI 393
Db 301 EPKDDIITHLSQOEKELKRIQIDSSDFLSTEEKEFLKKIQIDIRDSLSEEEKELMRI 360
QY 394 QVDSNPLSEKEKEFLKKLQIDOPYDINORLODPTGGLIDSPSINLDVRYKQYKRDIONID 453
Db 361 QVDSNPLSEKEKEFLKKLQIDOPYDINORLODPTGGLIDSPSINLDVRYKQYKRDIONID 420
QY 454 ALLHOSIGSTLYNKIYLYENNINNL7ATLQADLVSDTNTKINRGITNEKKNEKYSIS 513
Db 421 ALLHOSIGSTLYNKIYLYENNINNL7ATLQADLVSDTNTKINRGITNEKKNEKYSIS 480
QY 514 SNMYIVDINERPALDNEELKRIQISPTTRAGYLENGKLLIQRNIGLEIKVOIITKQSEK 573
Db 481 SNMYIVDINERPALDNEELKRIQISPTTRAGYLENGKLLIQRNIGLEIKVOIITKQSEK 540
QY 574 EYIRIDAVVVKSKIDTQIOEAQNLINQEWKALGLPYTKLITFNVNRRASNVESAY 633
Db 541 EYIRIDAVVVKSKIDTQIOEAQNLINQEWKALGLPYTKLITFNVNRRASNVESAY 600
QY 634 LILNEMKNNIOSDLIKKTYNTLVDSNGRFVFTDITLPIAEOYTHODEIYEQVHSGLYV 693
Db 601 LILNEMKNNIOSDLIKKTYNTLVDSNGRFVFTDITLPIAEOYTHODEIYEQVHSGLYV 660
```

Db 601 L11ENKNNI0SDLLIKKVTNYLVGNGRFVFTDITLPIAEOYTHODEIYEQVHSGLY 660
QY 694 PESRSTLLHGPSKGVGLRNDSEGFIEHFGHADVADYGYLLDKN0SLVYNSKKFIDIFKE 753
Db 661 PESRSTLLHGPSKGVGLRNDSEGFIEHFGHADVADYGYLLDKN0SLVYNSKKFIDIFKE 720
QY 754 EGSNLTSGRTNEAEFFAEAFRLMH 778
Db 721 EGSNLTSGRTNEAEFFAEAFRLMH 745

RESULT 3
PCT-US94-01624-2
; Sequence 2, Application PC/TUS9401624
; GENERAL INFORMATION:
; APPLICANT: Leppia, Stephen H.
; APPLICANT: Kilmpel, Kurt R.
; APPLICANT: Aroza, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREM
; STREET: Steuart Street Tower, 20th Floor, One Market
; STREET: Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01624
; FILING DATE: June 25, 1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31, 677
; REFERENCE/DOCKET NUMBER: 15280-115
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 776 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-01624-2

Query Match 95.9%; Score 3825; DB 5; Length 776;
Best Local Similarity 100.0%; Pred. No. 3.5e-239;
Matches 745; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 181 SDSGODLLFTNOLKEHPTDFSVFLEQNSNEQVFAFAFAYIIPQRHVDYQLTAPPA 240
QY 274 ENYMDKFNEQEIINLSLEELKDQRMLSRYEKWEKIKOHYHMSDSLSEGRGLLKQIPI 333
Db 241 FNYMDKFNEQEIINLSLEELKDQRMLSRYEKWEKIKOHYHMSDSLSEGRGLLKQIPI 300
QY 334 EPRKDDIINLSQEEKELKRIODSSDPLSTEEKFELKQIODIDSDSSEEEKELNRI 393
Db 301 EPRKDDIINLSQEEKELKRIODSSDPLSTEEKFELKQIODIDSDSSEEEKELNRI 360
QY 394 QVDSNPLSEKEKEFLKLLKLDIQPYDINORLODTGGLDPSINLDVROKQYRDIONID 453
Db 361 QVDSNPLSEKEKEFLKLLKLDIQPYDINORLODTGGLDPSINLDVROKQYRDIONID 420
QY 454 ALHQSIGSTLYNKKIYLYENMNINNLATLGLADLVSTDTKINRGIFFNEFKKRFYSIS 513
Db 421 ALHQSIGSTLYNKKIYLYENMNINNLATLGLADLVSTDTKINRGIFFNEFKKRFYSIS 480
QY 514 SNMYIVDINERPALDNBRKWRIOISPDRPAGYLENGKLILOINIGLEIKDYOIIOSEK 573
Db 481 SNMYIVDINERPALDNBRKWRIOISPDRPAGYLENGKLILOINIGLEIKDYOIIOSEK 540
QY 574 EYIRIDAKVYPKSKIDTKIQEALNINOENKALGLPKYTKLITFNVHNRVANSIVESAY 633
Db 541 EYIRIDAKVYPKSKIDTKIQEALNINOENKALGLPKYTKLITFNVHNRVANSIVESAY 600
QY 634 L11ENKNNI0SDLLIKKVTNYLVGNGRFVFTDITLPIAEOYTHODEIYEQVHSGLY 693
Db 601 L11ENKNNI0SDLLIKKVTNYLVGNGRFVFTDITLPIAEOYTHODEIYEQVHSGLY 660
QY 694 PESRSTLLHGPSKGVGLRNDSEGFIEHFGHADVADYGYLLDKN0SLVYNSKKFIDIFKE 753
Db 661 PESRSTLLHGPSKGVGLRNDSEGFIEHFGHADVADYGYLLDKN0SLVYNSKKFIDIFKE 720
QY 754 EGSNLTSGRTNEAEFFAEAFRLMH 778
Db 721 EGSNLTSGRTNEAEFFAEAFRLMH 745

RESULT 4
US-08-021-601-8
; Sequence 8, Application US/08021601
; Patent No. 5591631
; GENERAL INFORMATION:
; APPLICANT: Leppia, Stephen H.
; APPLICANT: Kilmpel, Kurt R.
; APPLICANT: Aroza, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,601
; FILING DATE: 19930212
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36, 016
; REFERENCE/DOCKET NUMBER: 1414.057

TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-8

Query Match 32.8%; Score 1309; DB 1; Length 472;
Best Local Similarity 98.8%; Pred. No. 4.1e-77;
Matches 255; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 34 AGGAGDGMHVKKEKKDKENKRDENRNTQOEHLKEIMKHIVKIGEAAYKKAEE 93
D 4 AGGAGDGMHVKKEKKDKENKRDENRNTQOEHLKEIMKHIVKIGEAAYKKAEE 63
QY 94 KLEKVPDVLVEMKKAIGKIYIVDGDITKHSLEALSEDKKKIKIDYKDALHHEHYV 153
D 64 KLEKVPDVLVEMKKAIGKIYIVDGDITKHSLEALSEDKKKIKIDYKDALHHEHYV 123
QY 154 AKEGEYVLVIOSSDEVVENTERKALNYYEIGKILSDILSKINOPYOKFLDVLNTIKNA 213
D 124 AKEGEYVLVIOSSDEVVENTERKALNYYEIGKILSDILSKINOPYOKFLDVLNTIKNA 183
QY 214 SDSGODLLFTNOLKEHPDTSVEFLQNSNEVOEFAKAFAYIEPQHNDVQLVAPEA 273
D 184 SDSGODLLFTNOLKEHPDTSVEFLQNSNEVOEFAKAFAYIEPQHNDVQLVAPEA 243
QY 274 FNYMDKFNEDQINLSLEE 291
D 244 FNYMDKFNEDQINLSLEE 261

RESULT 5
US-08-082-849B-8
Sequence 8, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppia, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-8

Query Match 32.8%; Score 1309; DB 1; Length 472;
Best Local Similarity 98.8%; Pred. No. 4.1e-77;
Matches 255; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 34 AGGAGDGMHVKKEKKDKENKRDENRNTQOEHLKEIMKHIVKIGEAAYKKAEE 93
D 4 AGGAGDGMHVKKEKKDKENKRDENRNTQOEHLKEIMKHIVKIGEAAYKKAEE 63
QY 94 KLEKVPDVLVEMKKAIGKIYIVDGDITKHSLEALSEDKKKIKIDYKDALHHEHYV 153
D 64 KLEKVPDVLVEMKKAIGKIYIVDGDITKHSLEALSEDKKKIKIDYKDALHHEHYV 123
QY 154 AKEGEYVLVIOSSDEVVENTERKALNYYEIGKILSDILSKINOPYOKFLDVLNTIKNA 213
D 124 AKEGEYVLVIOSSDEVVENTERKALNYYEIGKILSDILSKINOPYOKFLDVLNTIKNA 183
QY 214 SDSGODLLFTNOLKEHPDTSVEFLQNSNEVOEFAKAFAYIEPQHNDVQLVAPEA 273
D 184 SDSGODLLFTNOLKEHPDTSVEFLQNSNEVOEFAKAFAYIEPQHNDVQLVAPEA 243
QY 274 FNYMDKFNEDQINLSLEE 291
D 244 FNYMDKFNEDQINLSLEE 261

RESULT 6
PCT-US94-01624-8
Sequence 8, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppia, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: Stuart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-5043
TELEFAX: (415) 543-5043

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-8

Query Match 32.8%; Score 1309; DB 5; Length 472;
Best Local Similarity 98.8%; Pred. No. 4,1e-77;
Matches 255; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 34 AGCHGVGNHVKKEKKNKDEERNKTOEHLKEIMKHIVKTEVGEAVKKEAAE 93
|||||
DB 4 AGCHGVGNHVKKEKKNKDEERNKTOEHLKEIMKHIVKTEVGEAVKKEAAE 63
OY 94 KLEKVPDYLEMYKAIGGIYIVDGIKHSLEALSDDKKIKINDYKDALHHEHYV 153
|||||
DB 64 KLEKVPDYLEMYKAIGGIYIVDGIKHSLEALSDDKKIKINDYKDALHHEHYV 123
OY 154 AKEGYEPVLIQSSSEYVENTEKALNVYYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 213
|||||
DB 124 AKEGYEPVLIQSSSEYVENTEKALNVYYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 183
OY 214 SPSDGDLLFTNQLKHPDVSVEFLQNSNEVQVFAKAFAYIIEPQHRDVLQLYAPEA 273
|||||
DB 184 SPSDGDLLFTNQLKHPDVSVEFLQNSNEVQVFAKAFAYIIEPQHRDVLQLYAPEA 243
OY 274 FNYMDKFNEQEIINLSLE 291
|||||
DB 244 FNYMDKFNEQEIINLSLE 261

RESULT 7

US-08-021-601-10
Sequence 10, Application US/08021601
Patent No. 5591631

GENERAL INFORMATION:

APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414,057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-9880
TELEFAX: 404/688-0770

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:
LENGTH: 508 amino acids

TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-10

Query Match 32.8%; Score 1307; DB 1; Length 508;
Best Local Similarity 99.6%; Pred. No. 6,1e-77;
Matches 254; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 34 AGCHGVGNHVKKEKKNKDEERNKTOEHLKEIMKHIVKTEVGEAVKKEAAE 93
|||||
DB 1 AGCHGVGNHVKKEKKNKDEERNKTOEHLKEIMKHIVKTEVGEAVKKEAAE 60
OY 94 KLEKVPDYLEMYKAIGGIYIVDGIKHSLEALSDDKKIKINDYKDALHHEHYV 153
|||||
DB 61 KLEKVPDYLEMYKAIGGIYIVDGIKHSLEALSDDKKIKINDYKDALHHEHYV 120
OY 154 AKEGYEPVLIQSSSEYVENTEKALNVYYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 213
|||||
DB 121 AKEGYEPVLIQSSSEYVENTEKALNVYYEIGKILSRDILSKINOPYOKFLDVLNTIKNA 180
OY 214 SPSDGDLLFTNQLKHPDVSVEFLQNSNEVQVFAKAFAYIIEPQHRDVLQLYAPEA 273
|||||
DB 181 SPSDGDLLFTNQLKHPDVSVEFLQNSNEVQVFAKAFAYIIEPQHRDVLQLYAPEA 240
OY 274 FNYMDKFNEQEIINLS 288
|||||
DB 241 FNYMDKFNEQEIINLS 255

RESULT 8

US-08-082-849B-10
Sequence 10, Application US/08082849B
Patent No. 5677274

GENERAL INFORMATION:

APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993

CLASSIFICATION: 514

PRIOR APPLICATION NUMBER: US 08/021,601

FILING DATE: 12-FEB-1993

ATTORNEY/AGENT INFORMATION:

NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:
LENGTH: 508 amino acids

```

;      TYPE: amino acid
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
US-08-082-849B-10

```

Query Match	32.8%	Score 1307	DB 1	Length 508	
Best Local Similarity	99.6%	Pred. No. 6,1e-77			
Matches 254	Conservative 1	Mismatches 0	Indels 0	Gaps 0	
Qy	34	AGGHDGVMHVKKEKKNKDNKRKDEERNKTOEHLKEIKMHIIVKIEVGGEAVYKKAEE	93		
Db	1	AGGGDDVMHMKKEKKNNDENRKNKEERNKTOEHLKEIKMHIIVKIEVGGEAVYKKAEE	60		
Qy	94	KLLEKVPSPDVELEMYAIGKITYIVGGDITTKHLSLSPSKKIKINDYKGDLLHEHHY	153		
Db	61	KLLEKVPSPDVELEMYAIGKITYIVGGDITTKHLSLSPSKKIKINDYKGDLLHEHHY	120		
Qy	154	AKEGTEPVLVIQSSDDYVENTEKALNVYIEIGKILSRDILSKINPQYKFLDVLNTIKNA	213		
Db	121	AKEGTEPVLVIQSSDDYVENTEKALNVYIEIGKILSRDILSKINPQYKFLDVLNTIKNA	180		
Qy	214	SDSDGGODLLFNQLKEHPTDVSFEFLQONSNSNVQEVPAKAFYIIEPQHRVDLQLYAPPA	273		
Db	181	SDSDGGODLLFNQLKEHPTDVSFEFLQONSNSNVQEVPAKAFYIIEPQHRVDLQLYAPPA	240		
Qy	274	FNHYMDKFNQEOELINS 288			
Db	241	FNHYMDKFNQEOELINT 255			

```

RESULT      9
PCT-US94-01624-10
: Sequence 10. Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: Steuart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 508 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

```

		PCT-US94-01624-10	
	Query Match	32.8%	Score 1307; DB 5; Length 508;
	Best Local Similarity	99.6%	Pred. No. 6.1e-77;
	Matches 254; Conservative	1; Mismatches	0; Indels 0; Gaps 0;
QY	34 AGGHDDVGMNHYKEREKKNNDENKRDDEENKQOEHLKLEIMKHIVKIEVKGEEVAKKEAAE	93	
Db	1 AGGHDDVGMNHYKEREKKNNDENKRDDEENKQOEHLKLEIMKHIVKIEVKGEEVAKKEAAE	60	
QY	94 KLEKVPSPDLVEMXKAIGKIYIVDGDITKHSLEALSDEKKRKIDYKDALHEHYV	153	
Db	61 KLEKVPSPDLVEMXKAIGKIYIVDGDITKHSLEALSDEKKRKIDYKDALHEHYV	120	
QY	154 AKEGEPVLVIQSSSDVYENTEKALNVYVEIGKIISRLISKTNPYOKFLDVLNTIKNA	213	
Db	121 AKEGEPVLVIQSSSDVYENTEKALNVYVEIGKIISRLISKTNPYOKFLDVLNTIKNA	180	
QY	214 SDSDGQDLLFTNQLKEHPTDSVEFLLEQNSNEVGEVFAKAFAYUYIEPQHRDVLQIYARPA	273	
Db	181 SDSDGQDLLFTNQLKEHPTDSVEFLLEQNSNEVGEVFAKAFAYUYIEPQHRDVLQIYARPA	240	
QY	274 FNYMDKENEQEIINLS 288		
Db	241 FNYMDKENEQEIINLT 255		

```

: RESULT 10
: US-08-021-601-6
: Sequence 6, Application US/08021601
: Patent No. 5591631
: GENERAL INFORMATION:
: APPLICANT: Leppia, Stephen H.
: APPLICANT: Klimpel, Kurt R.
: APPLICANT: Nichols, Peter J.
: APPLICANT: Aroa, Naveen
: APPLICANT: Singh, Yogendra
: TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
: TITLE OF INVENTION: RELATED METHODS
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Needle & Rosenberg, P.C.
: STREET: 133 Carnegie Way, Suite 400
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: USA
: ZIP: 30303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/021,601
: FILING DATE: 19930212
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Spratt, Gwendolyn D.
: REGISTRATION NUMBER: 36,016
: REFERENCE/DOCKET NUMBER: 1414,057
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 404/688-0770
: TELEFAX: 404/688-9880
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 456 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-021-601-6

```


Query Match 32.8%; Score 1306; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 6.2e-77;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGCHGVGMHVKERKKNKDEBNKTOEHLKEIKKHIVKIEVGEEAVKKEAAE 93
DB 1 AGCHGVGMHVKERKKNKDEBNKTOEHLKEIKKHIVKIEVGEEAVKKEAAE 60
QY 94 KLEKPSVLEMYKKAIGKITYVDGDIKHSLEALSEDKKKIKDIYGGDALLHEHYV 153
DB 61 KLEKPSVLEMYKKAIGKITYVDGDIKHSLEALSEDKKKIKDIYGGDALLHEHYV 120
QY 154 AKEGYEVLTIOSESEYVENTEKALNVYIEIGKILSRDISKINOPYOKFLDVLNTIKNA 213
DB 121 AKEGYEVLTIOSESEYVENTEKALNVYIEIGKILSRDISKINOPYOKFLDVLNTIKNA 180
QY 214 SDSGDGLFTNOLKEHPDTSVEFLQNSNEVOEYFAKAFAYIIEPQHRDVLQLYAPEA 273
DB 181 SDSGDGLFTNOLKEHPDTSVEFLQNSNEVOEYFAKAFAYIIEPQHRDVLQLYAPEA 240
QY 274 FNYMDKFNQEOEINL 287
DB 241 FNYMDKFNQEOEINL 254

RESULT 11

US-08-082-849B-6
Sequence 6, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-6

Query Match 32.8%; Score 1306; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 6.2e-77;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGCHGVGMHVKERKKNKDEBNKTOEHLKEIKKHIVKIEVGEEAVKKEAAE 93
DB 1 AGCHGVGMHVKERKKNKDEBNKTOEHLKEIKKHIVKIEVGEEAVKKEAAE 60
QY 94 KLEKPSVLEMYKKAIGKITYVDGDIKHSLEALSEDKKKIKDIYGGDALLHEHYV 153
DB 61 KLEKPSVLEMYKKAIGKITYVDGDIKHSLEALSEDKKKIKDIYGGDALLHEHYV 120
QY 154 AKEGYEVLTIOSESEYVENTEKALNVYIEIGKILSRDISKINOPYOKFLDVLNTIKNA 213
DB 121 AKEGYEVLTIOSESEYVENTEKALNVYIEIGKILSRDISKINOPYOKFLDVLNTIKNA 180
QY 214 SDSGDGLFTNOLKEHPDTSVEFLQNSNEVOEYFAKAFAYIIEPQHRDVLQLYAPEA 273
DB 181 SDSGDGLFTNOLKEHPDTSVEFLQNSNEVOEYFAKAFAYIIEPQHRDVLQLYAPEA 240
QY 274 FNYMDKFNQEOEINL 287
DB 241 FNYMDKFNQEOEINL 254

RESULT 12

PCT-US94-01624-6
Sequence 6, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: Steuart Street Tower, 20th Floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-6

Query Match 32.8%; Score 1306; DB 5; Length 456;
Best Local Similarity 100.0%; Pred. No. 6.2e-77;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 AGHGVDGMHVKKEKKNDEKRRKQOEHLKEIMKHIVKEVGEAAVKEAAE 93
D 1 AGHGVDGMHVKKEKKNDEKRRKQOEHLKEIMKHIVKEVGEAAVKEAAE 60
QY 94 KLLKEVPDVLDMYKATGKLYIVDGDITKHSLEALSEDKKIKDIYKQALLHEHYV 153
D 61 KLLKEVPDVLDMYKATGKLYIVDGDITKHSLEALSEDKKIKDIYKQALLHEHYV 120
QY 154 AKEGEPLVLIOSSEDDYVENTEKALNVEYEIGKILSRDLSKINOPYOKFLDVLNTIKNA 213
D 121 AKEGEPLVLIOSSEDDYVENTEKALNVEYEIGKILSRDLSKINOPYOKFLDVLNTIKNA 180
QY 214 SDSGODLFTNOLKEHPTDSVEFLQNSNEVOEVPKAFAYITIEPOHVDLYAPEA 273
D 181 SDSGODLFTNOLKEHPTDSVEFLQNSNEVOEVPKAFAYITIEPOHVDLYAPEA 240
QY 274 FNYMDKFEQELN 287
D 241 FNYMDKFEQELN 254
RESULT 13
5183745-3
PATENT NO. 5183745
APPLICANT: DANCHIN, ANTOINE; GLASER, PHILIPPE; KRIN, EVELYN;
BARZU, OCTAVIEN; LADANT, DANIEL; ULLMAN, AGNES
TITLE OF INVENTION: ADENYL CYCLASE DERIVATIVES AND THEIR
BIOLOGICAL USES
NUMBER OF SEQUENCES: 13
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/426,541
FILING DATE: 25-OCT-1989
SEQ ID NO: 3
LENGTH: 800
5183745-3
Query Match 12.6%; Score 501; DB 6; Length 800;
Best Local Similarity 24.6%; Pred. No. 1,1e-24;
Matches 202; Conservative 145; Mismatches 281; Indels 194; Gaps 36;
QY 4 KKEFT---KVISMCATATLTSQPVPIPLVQAGHGVDGMHVKKEKKNDEKRRDE 59
D 3 RNRKIPKRFSTIISVLLFAISSSOAIEVNA-----MNEHYTESDIKRNKHT 49
QY 60 ERNKTQEEHLKEIMKHIVKEVGEAAVKEAAEKLKEVPDVLDMYKATGKLYIVDG 119
D 50 EKNKTEKFKRDSINNVLKTEFTNETLTKIQOTDLKKIPKVDLEYISEIGETIYFDI 109
QY 120 DITHHISLEALSEDKKIKDIYKQALLHEHYVAKEGEPLVLIOSSEDDYVENTEKALN 179
D 110 DLVHKEKELQDISEEKKSMNSRGEKVPAPSRFVEKKRETFLKI-NKDAINSEOSKE 168
QY 180 VYEIGKILSRDLSKINOPYOKFLD--VLNTIKNA--SDSGODLFTNOLKE---HPT 232
D 169 VYEIGKISLDITISK-----DKSLDPEFLNLSLSDSSDLSLQKREKLELNK 223
QY 233 DVSVEFLQNSNEVOEVPKAFAYITIEPOHVDLYAPEAFTNMDKFEQELNLSREL 292
D 224 SIDINFIKENLTERQHFSLAFSYFAPDRHTVELYAPMEFVYNNKL----- 271
QY 293 KDQKMLSTYER--WEKIKOHVHMSDGLSEGRGLLKLOIPIEPKDDIHSLSQEEKE 350
D 272 -----EKGEPEKI-----SESLKKEG-----VEKDRIDVL---KGEKA 301
QY 351 LKRIQDSOFLSTEEKEFLKLIQIDIRDSLSEEEKELNRIQVDSNPLSEKKEFLK 410
D 302 L-----KASGLVPEHNAFAFKI-----ARLNTYILFRPVNKLATN---LI 339
QY 411 KLLKDIOPYDINORLQDTGGLDPSINLDVAKOY-----KRDIONIDL--HQ--SIG 461
D 340 KSGVATKGLANHGKSDMGVAGYIIPFDQDLSTKKGQGLAVKEKGLNKKSTITEHIGIG 399

QY 462 STLNNKIYL-YENMNINNLATLGLADLVSDTNKINRGIFNEKKNK-----YSISSN 515
D 400 -----KIPLKIDHRIEELK-----ENGIIKQKKEIDNKKYLLLESN 438
QY 516 YMIYDINERPALDMERLKWRIQOLSPDRAGYLE--NGKILIORNIGLEKQVQ--IIKSEK 573
D 439 NOVVEF--RIDENNEVOYK-----TKEGKITVGEFPMRNIEVAKNVEGYLPLTA 490
QY 574 EYIRDAKVVPSKIDYFIOEAOQLINQENMKALGLPKYTK-----LITFVNHRYA 625
D 491 DY---DLFALPRLSLEIKQIIPQ---KEMDKVNTPLSLEKQGVNLLIKYIERK-- 541
QY 626 SNIVESAYLLINEKNKNIQSDLIKVTNYLVQNGREYFDTILPNAIOYTHOD--EY 683
D 542 ---PDSTKGLTSLNOKQMLDRL-----NEAVKYTGTCGDVNVNHTGQDNNEFP 587
QY 684 EOVSKGIVPESRSILHG---PSKGVELRNDSEGFHEGHAVDYA---GYL---- 732
D 588 EKDNIEFIINEGEFILTAKNEMTGRFIEKNITGKDYLYFNRSYNNIAPGNKAYIEWTD 647
QY 733 -LDRKNSDLVTSKKFIDIFKEGSLNLTSGRTNEAEFFAA 773
D 648 PITKAKINTIFTSAEFT-----KNLSIRSSNNGVYKDS 682

RESULT 14
US-08-973-462-8
Sequence 8, Application US/08973462B
PATENT NO. 6191270
GENERAL INFORMATION:
APPLICANT: DRUTHE, PIERRE
TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
FILE REFERENCE: 0660-0125-0 PCT
CURRENT APPLICATION NUMBER: US/08/973,462B
CURRENT FILING DATE: 1998-02-06
EARLIER APPLICATION NUMBER: PCT/FR96/00894
EARLIER FILING DATE: 1996-06-12
EARLIER APPLICATION NUMBER: FR 95/07007
EARLIER FILING DATE: 1995-06-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 1786
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

Query Match 5.7%; Score 228.5; DB 4; Length 1786;
Best Local Similarity 18.9%; Pred. No. 1,2e-06;
Matches 155; Conservative 134; Mismatches 271; Indels 259; Gaps 32;
QY 44 VKEKEKKNDEKRRKQOEHLKEIMKHIVKEVGEAAVKEAAEKLKEVP--- 100
D 891 IEEKLEELHEVNLVALENTQSEEEKKEVIDVI-----EEVKEEVAATLITVEQAE 942
QY 101 -----SDVLEMYKATGKIY--IYDG--DIRKHSLEALSE--- 132
D 943 EKSANTTIEIPENLEENAVESNEVNAEMLKNETFTVFTYLDKVEYVEISGESLENEM 1002
QY 133 DKRKIKDIYGDALLHEHYVAK--EGEPLVLIOSSE-----DYVENTEKAL 178
D 1003 DKAFSEIFDVNVKQIENLLGMFRSIFSTVIOSEKENVVDNENVSSLDINEMKKEGL 1062
QY 179 -----NVYEIGKILSRD---ILSKINOPYOKFLDVLNTIK 211
D 1063 LNKLENISSTEGVOETVTEHVEQNVVYVDVPAKMDQFLGLINAGAGLKEMFNLEDFVK 1122
QY 212 NASDSODGDLFTNOLKEHPTDSVEFLQNSNEVOEVPKAFAYITIEPOHVDLYAP 271

Db 1123 SES-----DVYVEEIKDEP-----VQKEVEKETSIEEENIVDL EE 1163
QY 272 EARNYMDK-----NEQETNLSLEELKQRMLSREKMEKIK 308
Db 1164 EKEDLDKMDAVEESIEISDSKRETESIKDEKDVSLVEEYODMDSEVEKLEK 1223
QY 309 OHYOHV-----SDLSSEGRG-----LKKLOPIE-----PKD 338
Db 1224 NMEELMKDAVEINDITSKLIETOELEVEADLIKMEKLEKALSDESKETIDAKD 1283
QY 339 DIHLSQEEKEK-----LKRQIDS-----SDFLSTEE-----KEF----- 370
Db 1284 DLEKYLEEHDITTLDEVEVELKVDKEDKLEKVDLKDLEEDLIKVEKKEIKLESFLE 1343
QY 371 -----LKKLOIDRDSL-----EERKELNRIQVDSNPLSEKEFEKLKLDI 416
Db 1344 DYKELKTIEFDILEEKKEIEKDHFEKEFEAEIKDELDILEVSSLEVEEKEKLE-- 1400
QY 417 OPYDINORLOD-----TGGLIDSPSINLDVRKQYKRDIONDA 454
Db 1401 EHELEKEVEHIIISGAHIKGLEEDLEEVVDLKGSTLMDLKGDMELGDMKESLEEDVT 1460
QY 455 LHQSIGSTLYNKTYLKENNNINLTATGADLVSDTNTKINRG-----IFNEFKNF 508
Db 1461 KLGERVESLK-----DVLSSALGMDDEQMKTRKKAQRPKEVLEKVEYKEEP 1508
QY 509 KYSISSNYMIVDINERPAID-----NERLKWRIQ-----LSPTPRAGYLENGKLILORNI 558
Db 1509 KKKITTKKVFEDIKDEPEKDEIVEEMKDEIDIEDEEDIEEDKEDIEDIEDIEDI 1568
QY 559 GLEIKD--VOIIRKSEKEYIRIDAKVVPKSIDTKIOEAOQUNIOENKRALGLPKYTKLI 616
Db 1569 G-EDKREVIDLIYQEKREIKRYAK--KKLEKKVEEGVSGLKHAVEVW--KIVQKI 1621
QY 617 TENVHNRYASNIYESAYLLINEMKNNIQSDLIKRYNYL 655
Db 1622 DKREV-DKEVSKALESKNDVTNVLKON--QDFFSKVKNFV 1657

RESULT 15
US-08-480-604A-6
Sequence 6, Application US/08480604A
Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESSES:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-6

Query Match 5.1%; Score 202; DB 1; Length 2710;
Best Local Similarity 20.4%; Pred. No. 0.0001;
Matches 189; Conservative 133; Mismatches 264; Indels 340; Gaps 49;

QY 48 EKKNDENRRDEERKNTQOEHLKEI-MKHIVKIEVKG--EAVKKAERKLEK--V 99
Db 212 EYNDETVLSYRNS-----LKRKINSHGIDIRANSFTQOELLINYSOELLNRGMIAA 266
QY 100 PSDYLEM--YKALGKRYI-----VDGDTIKHS-----LEALSEKKK 136
Db 267 ASDIVRLALKNFGG-VYLLVMDLPGIHSDFKTSRPSISGLDRWMIKLEALIMKYKY 325
QY 137 IKDIYGRDALHEHYVAKGEYEP-----LVYOSSEDYVENTEKALN----- 179
Db 326 INN-----YTSNFPKLDQOLKDNFKLIIESKSEKEITSKLENLVSDLEI 372
QY 180 -VYIEIGKILSRDLSK-----INOPYQFLVDLNTIKNASDSGDGLFTNOLK- 228
Db 373 KIFALGSVINQALISKQGSYLTNLVIEQYKNRYQFLNQHLPALIESDNN--FTDTTKI 429
QY 229 EHPIDSEVEFLQNSNEVOEYAKAFAYIE-----POHRYLOLYAPEAF--NYMDKFE 282
Db 430 FHDLSLFNSATVENS-----MFLTKIAPYLOVGFMPARSTISLSGGAASAVYDFINL 483
QY 283 QETNL-----SLEELKQRMLSRYE-----KMEKIKOHYOHMSDS 317
Db 484 QENTIENTLAKSLIEKFPENNINSQLTEQFINSLMFDASAKYQEKYVRYDT--GGS 541
QY 318 LSEEG-----RGLKLIQIP-----IEPKKDI-----IHS 343
Db 542 LSEDNQVDFPNKNTALDKNYLLNNKIPSNVNEAGSKVYVYIIOQSDDISYEVATCNLFS 601
QY 344 LSOEKKELKRIQIDS--SDFLSTEEKEFLKLOIDRDSLSEEEKLLN----- 391
Db 602 KNPKNSTIIQIRNMNESKSYFLSDGSESLILNKRIPIERLKNEXKAVYFEGHGXDEFN 661
QY 392 -----RIVQSSNPLSEKEFEKLKLDIOP-----VDINORLODGTGLI 432
Db 662 TSEFARLSVDS--LSNEISFLDTIKLDISPKNVEYNLGCMMFSTDFNVEEYEPGKLI 718
QY 433 DS-----PSINLDV--RKQYKRDION--IDALH-----OSIGSTLYNKI 468
Db 719 LSIWDKITSITLPDVNKNKSIITGANQYEVRIINSEGRKELLASHGKWIKEAIAIMSDLSKE 778

```

QY 469 YLYENMNINLTATLGADLVSTDTNTKINRGIFNEFKKNFRYSISSNYM--IVDINERPA 526
Db 779 YIF-----FDSIDN-----KLKAKSKNIPGLASISE--- 804
QY 527 LDNERLKWRIOQSPDTRAGYIENGKLIQNRNIGLEIKDVQIIOSEKEYIRID---AKV 582
Db 805 -DIRTLLDASVSPDTKF-IINMLKLNIESSIG-----DIYIEKLEPYKN 848
QY 583 VPKSKIDTKIOEAOI--NINOEMNKALGLPKYTKLITFNVHNRVASNIVESAYLILNEMK 640
Db 849 IINHSDLDLIDEFNLEWVSD-----LYELKIL-----NNLDEKYLISFE-- 889
QY 641 NNIOSDLIKVNTYLV---DNGRFVFTDITLPNIAEQYTHODEIYEQVHSGLYPES 696
Db 890 ----DISKNNSTYSVRFINKSNNGESYVE--TEKEIFSKYS--EHITKEIST----- 933
QY 697 RSIILHGFSKGVLELRNDESEGFIERGHAVDDYAGYLLDKNOSDLVT-----NSKKFTIDIF 751
Db 934 -----IKNS-----ITDVNGNLDNIQILDHTSQVNTLNAAFIOSL 970
QY 752 KEEGSN---LTSYGRTEAEFFAEAF 774
Db 971 IDYSSNKDVNLDSLSTSVKVOQLXAOIF 996

```

Search completed: December 2, 2001, 13:49:24
 Job time: 204 sec